Cochrane Library

Cochrane Database of Systematic Reviews

Interventions to improve return to work in depressed people (Review)

Nieuwenhuijsen K, Faber B, Verbeek JH, Neumeyer-Gromen A, Hees HL, Verhoeven AC, van der Feltz-Cornelis CM, Bültmann U

Nieuwenhuijsen K, Faber B, Verbeek JH, Neumeyer-Gromen A, Hees HL, Verhoeven AC, van der Feltz-Cornelis CM, Bültmann U.
Interventions to improve return to work in depressed people.
Cochrane Database of Systematic Reviews 2014, Issue 12. Art. No.: CD006237.
DOI: 10.1002/14651858.CD006237.pub3.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Interventions to improve return to work in depressed people

Karen Nieuwenhuijsen[1], Babs Faber[2], Jos H Verbeek[3], Angela Neumeyer-Gromen[4], Hiske L Hees[5], Arco C Verhoeven[6], Christina M van der Feltz-Cornelis[7], Ute Bültmann[8]

[1]Coronel Institute of Occupational Health/Dutch Research Center for Insurance Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands. [2]Coronel Institute of Occupational Health/Dutch Research Center for Insurance Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands. [3]Cochrane Occupational Safety and Health Review Group, Finnish Institute of Occupational Health, Kuopio, Finland. [4]Deutsche Krankenhausgesellschaft (The German Hospital Federation), Berlin, Germany. [5]Program for Mood Disorders, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands. [6]NVAB Centre of Excellence, Netherlands Society of Occupational Medicine, Utrecht, Netherlands. [7]Department Tranzo, Faculty of Social and Behavioural Sciences, Tilburg University, Tilburg, Netherlands. [8]Department of Health Sciences, Community and Occupational Medicine, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Contact address: Karen Nieuwenhuijsen, Coronel Institute of Occupational Health/Dutch Research Center for Insurance Medicine, Academic Medical Center, University of Amsterdam, PO Box 22700, Amsterdam, 1100 DE, Netherlands. K.Nieuwenhuijsen@amc.nl.

Editorial group: Cochrane Work Group.
Publication status and date: Edited (conclusions changed), published in Issue 12, 2014.

Citation: Nieuwenhuijsen K, Faber B, Verbeek JH, Neumeyer-Gromen A, Hees HL, Verhoeven AC, van der Feltz-Cornelis CM, Bültmann U. Interventions to improve return to work in depressed people. *Cochrane Database of Systematic Reviews* 2014, Issue 12. Art. No.: CD006237. DOI: 10.1002/14651858.CD006237.pub3.

ABSTRACT

Background

Work disability such as sickness absence is common in people with depression.

Objectives

To evaluate the effectiveness of interventions aimed at reducing work disability in employees with depressive disorders.

Search methods

We searched CENTRAL (The Cochrane Library), MEDLINE, EMBASE, CINAHL, and PsycINFO until January 2014.

Selection criteria

We included randomised controlled trials (RCTs) and cluster RCTs of work-directed and clinical interventions for depressed people that included sickness absence as an outcome.

Data collection and analysis

Two authors independently extracted the data and assessed trial quality. We used standardised mean differences (SMDs) with 95% confidence intervals (CIs) to pool study results in the studies we judged to be sufficiently similar. We used GRADE to rate the quality of the evidence.

Main results

We included 23 studies with 26 study arms, involving 5996 participants with either a major depressive disorder or a high level of depressive symptoms. We judged 14 studies to have a high risk of bias and nine to have a low risk of bias.

Work-directed interventions

We identified five work-directed interventions. There was moderate quality evidence that a work-directed intervention added to a clinical intervention reduced sickness absence (SMD -0.40; 95% CI -0.66 to -0.14; 3 studies) compared to a clinical intervention alone.

There was moderate quality evidence based on a single study that enhancing the clinical care in addition to regular work-directed care was not more effective than work-directed care alone (SMD -0.14; 95% CI -0.49 to 0.21).

There was very low quality evidence based on one study that regular care by occupational physicians that was enhanced with an exposure-based return to work program did not reduce sickness absence compared to regular care by occupational physicians (non-significant finding: SMD 0.45; 95% CI -0.00 to 0.91).

Clinical interventions, antidepressant medication

Three studies compared the effectiveness of selective serotonin reuptake inhibitor (SSRI) to selective norepinephrine reuptake inhibitor (SNRI) medication on reducing sickness absence and yielded highly inconsistent results.

Clinical interventions, psychological

We found moderate quality evidence based on three studies that telephone or online cognitive behavioural therapy was more effective in reducing sick leave than usual primary or occupational care (SMD -0.23; 95% CI -0.45 to -0.01).

Clinical interventions, psychological combined with antidepressant medication

We found low quality evidence based on two studies that enhanced primary care did not substantially decrease sickness absence in the medium term (4 to 12 months) (SMD -0.02; 95% CI -0.15 to 0.12). A third study found no substantial effect on sickness absence in favour of this intervention in the long term (24 months).

We found high quality evidence, based on one study, that a structured telephone outreach and care management program was more effective in reducing sickness absence than usual care (SMD - 0.21; 95% CI -0.37 to -0.05).

Clinical interventions, exercise

We found low quality evidence based on one study that supervised strength exercise reduced sickness absence compared to relaxation (SMD -1.11; 95% CI -1.68 to -0.54). We found moderate quality evidence based on two studies that aerobic exercise was no more effective in reducing sickness absence than relaxation or stretching (SMD -0.06; 95% CI -0.36 to 0.24).

Authors' conclusions

We found moderate quality evidence that adding a work-directed intervention to a clinical intervention reduced the number of days on sick leave compared to a clinical intervention alone. We also found moderate quality evidence that enhancing primary or occupational care with cognitive behavioural therapy reduced sick leave compared to the usual care. A structured telephone outreach and care management program that included medication reduced sickness absence compared to usual care. However, enhancing primary care with a quality improvement program did not have a considerable effect on sickness absence. There was no evidence of a difference in effect on sickness absence of one antidepressant medication compared to another. More studies are needed on work-directed interventions. Clinical intervention studies should also include work outcomes to increase our knowledge on reducing sickness absence in depressed workers.

PLAIN LANGUAGE SUMMARY

Interventions to help depressed people resume work

Depression is a major problem that affects about 300 million people globally. Symptoms of depression include the core symptoms of low mood or loss of interest coupled with other symptoms such as feelings of inadequacy and hopelessness or sleep problems. These symptoms usually impair functioning and therefore sickness absence is common in people with depression. We evaluated the effectiveness of interventions that can help depressed workers to resume work activities.

Studies we found

We found 23 studies, involving 5996 participants, that looked at the effects on sick leave of changes at work that were in addition to regular treatment, better psychological treatment, improving primary care, antidepressant pills and exercise.

Effects of changes at work in addition to regular care

In three studies with 251 participants, researchers looked at changes at work such as work modification or coaching in addition to regular care and found that these reduced sickness absence to a moderate extent.

In two studies, researchers tried to improve care that was already directed at changes at work but did not find any effects of these improvements on sick leave.

Effects of psychological treatment

In three studies with 326 participants, researchers found that cognitive behavioural therapy that was provided online or by telephone reduced sickness absence to a moderate extent compared to regular care.

In one high quality study, a special care programme carried out via the workplace also reduced sick leave when compared to regular care.

Effects of antidepressant pills

Three studies compared antidepressant pills with each other but there were no consistent effects on sickness absence.

Improving primary care

Improving primary care through quality improvement programs for general practitioners did not reduce sickness absence in three studies.

Exercise

One study found that participants had a reduction in sick leave after doing stretching exercises. Two other studies did not find an effect on sick leave after physical exercises such as running or using the gymnasium.

More studies should look at the effects of changes at work. Regular clinical studies should also measure the effects on sick leave because this is an important consequence of depression.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON [Explanation]

Work-directed intervention plus clinical intervention compared to clinical intervention alone for depressive disorder

Patient or population: Persons with depressive disorder
Settings: Two studies were conducted in outpatient and one in a workplace or Employee Assistance Program
Intervention: Work-directed intervention plus clinical intervention
Comparison: Clinical intervention alone

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Clinical intervention	Work-directed intervention plus clinical intervention				
Days of sickness absence Follow up: 4 - 12 months		The mean days of sickness absence in the intervention groups was **0.4 standard deviations lower** (0.66 to 0.14 lower)	SMD -0.4 (-0.66 to -0.14)	251 (3 studies)	⊕⊕⊕◯ **moderate**[1]	A standard deviation of 0.5 represents a moderate difference between groups

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **SMD**: Standardised Mean Difference

Intervention description:
In two studies, an occupational therapy program focusing on work reintegration, combining modified work and supportive interventions was added to clinical care. In one study a regular Employee Assistance program was expanded and incorporated work coaching and modification
GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Downgraded one level because N < 400

BACKGROUND

Description of the condition

Depression is a major public health problem, with 298 million cases of major depressive disorders at any time point in 2010 (Ferrari 2013). The worldwide point prevalences of depressive disorder were 4.4% in both 2005 and 2010 (Ferrari 2013). Symptoms of depressive disorder include the presence of one or two core symptoms of low mood and loss of interest, coupled with other symptoms such as feelings of inadequacy and hopelessness, sleep disturbance, weight change, fatigue, impaired concentration, agitation or slowing down of movement and thought, and suicidal ideation (APA 2013). Depressive disorders can be classified along a continuum by the levels of symptom severity, number of mental or physical symptoms, and duration. Corresponding diagnostic categories range from persistent depression (dysthymia) and subclinical states (minor depressive disorder) to major depressive disorder (APA 1994; APA 2013).

Besides the serious consequences in terms of individual suffering, depression has a large impact on social functioning and the ability of patients to work (Hirschfeld 2000; Lerner 2008). In a population of US workers, the 12-month prevalence of major depressive disorder was found to be 6% and was associated with 27.2 lost workdays per ill worker per year (Kessler 2006). In terms of annualised human capital loss to employers in the US labour force, this amounted to about USD 36 billion (Kessler 2009). The high prevalence of depressive disorders, combined with the impact on work disability, has extensive societal consequences. In 1990, major depressive disorders were the 15th leading contributor to the global burden of disease in terms of 'Disability Adjusted Life Years' (DALYs), which is the sum of years of productive life lost due to premature mortality and the years of productive life lost due to disability. Data from the global burden of disease study in 2010 showed that depressive disorders are now ranked 11th (Murray 2012).

While working is important from a societal point of view, work is also an important aspect of the quality of life of individuals (Bowling 1995). Work provides income, structure, and social interactions. One salient consequence of depression is absenteeism, but depression can also affect the at-work productivity for workers (Lerner 2008). Depressed workers experience specific limitations in their ability to function at work. These limitations include performing mental and interpersonal tasks (Adler 2006; Burton 2004). The quality of work performance can also be affected, as was shown in studies focusing on errors and safety issues (Haslam 2005; Suzuki 2004). Depressed workers may need to make an extra effort to be productive during their work (Dewa 2000), which may lead to spillover effects of fatigue after work.

Description of the intervention

Work ability of depressed workers can be targeted by interventions. First of all, work-directed interventions aim to ameliorate the consequences of the depressive disorder on the ability to work. These types of interventions either target the work itself, by modifying the job task, or (temporarily) reduce the working hours. Work-directed interventions can also support the worker in dealing with the consequences of their depression at the workplace.

Second, clinical interventions aimed at reducing depression symptoms may improve work ability (Hees 2013b). Current clinical practice guidelines for the treatment of major depressive disorder recommend pharmacotherapy, psychotherapy, or a combination of both (APA 2010; NICE 2010). Pharmacologic treatment for major depressive disorder includes antidepressant medication such as tricyclic antidepressants (TCAs), selective serotonin reuptake inhibitors (SSRIs), monoamine oxidase inhibitors (MAO inhibitors), and selective norepinephrine reuptake inhibitors (SNRIs). With regard to psychotherapy, cognitive behavioural therapy (CBT) and interpersonal therapy are considered effective treatment options (NICE 2010). Exercise has been increasingly used as an alternative to pharmacological or psychotherapeutic interventions (Cooney 2013).

How the intervention might work

Work-directed interventions are deemed to reduce work disability by creating a work environment better suited for a depressed worker, such as modifying work tasks or working hours. Moreover, the worker can be supported in dealing with the depression at work by a gradual return to work program or by enhancing skills to cope with work situations (Lagerveld 2012). Clinical interventions may reduce work disability by reducing depressive symptoms, thereby eliminating the obstacles to working.

Why it is important to do this review

Considering the impact of depressive disorders on the occupational health of many affected workers, it is vital to know what types of interventions are effective in improving occupational health. In the first version of this review, in 2008, we concluded that there was an urgent need to evaluate interventions that address work issues in future research. Since then, several such studies have been published underpinning the need for an update of the review.

OBJECTIVES

The goal of this review was to evaluate the effectiveness of interventions aimed at reducing work disability in employees with depressive disorders.

We considered the effectiveness of two types of interventions:

1. work-directed interventions, i.e. addressing the work or the work-worker interface as part of the clinical treatment or as a stand-alone intervention; and

2. clinical interventions, i.e. treatment of depressive disorder without a focus on work.

METHODS

Criteria for considering studies for this review

Types of studies

We included all randomised controlled trials (RCTs), including cluster RCTs, in this review. We did not use any language restrictions.

Types of participants

Patient characteristics and setting

The population was limited to adult (that is over 17 years old) workers (employees or self-employed). We included participants from occupational health settings, primary care, or outpatient care settings. We based the selection of the studies on the primary outcome only.

Diagnosis

We defined depressive disorder as a main diagnosis fulfilling the criteria of the Diagnostic and Statistical Manual (DSM-IV) (APA 1994; APA 2013), the Research Diagnostic Criteria (RDC) (Spitzer 1979), or the International Classification of Disease (ICD-10) (WHO 1992) for one of the following disorders: dysthymic disorder, minor depressive disorder, or major depressive disorder. We also included studies that defined depressive disorder as a level of depressive symptoms assessed by validated self-report instruments published in peer-reviewed journals. An examples is the Beck Depression Inventory (BDI) (Beck 1987); or clinician-rated instruments such as the Hamilton Depression Rating Scale (HDRS) (Hamilton 1967) or the Montgomery-Åsberg Depression Rating Scale (MADRS) (Montgomery 1979).

Exclusion criteria

We excluded studies involving workers with a primary diagnosis of a common mental disorder other than a depressive disorder. We did not exclude workers with a co-morbidity from other common mental disorders (such as anxiety disorders), but we did exclude workers with bipolar disorders or depressive disorders with psychotic features.

Types of interventions

We included all interventions aimed at reducing work disability, thereby differentiating work-directed interventions from clinical interventions. Examples of work-directed interventions are light duty, graded work exposure, or supportive interventions enhancing the coping of the worker with depression in the workplace. We categorised work-directed interventions as:

1. modified work, modified working hours or job tasks;
2. supportive, supporting the worker in coping with depression at the workplace; and
3. a combination of modified work and supportive interventions.

We divided clinical interventions into:

1. antidepressant medication, interventions that use any type of antidepressant medication at any dose;
2. psychological, where psychological interventions are restricted to cognitive behavioural interventions (CBT), interpersonal therapy (IPT), problem solving therapy (PST), psychodynamic therapy, counselling, and occupational therapy, undertaken by qualified trained therapists; and
3. physical, where physical interventions are restricted to those using exercise, i.e. strength or aerobic training.

Main comparisons

We conducted, where data were available, the following treatment comparisons in order to address the review's objectives.

1. Work-directed interventions

- Work-directed intervention plus clinical intervention versus clinical intervention alone
- Work-directed intervention plus clinical intervention versus work-directed alone
- Any work-directed intervention versus no intervention or care as usual
- Any work-directed intervention versus an alternative work-directed intervention

2. Clinical interventions, antidepressant medication

- Any antidepressant medication versus any other antidepressant medication
- Any antidepressant medication versus placebo

- Any antidepressant medication versus any psychological intervention

3. Clinical interventions, psychological

- Any psychological intervention versus other psychological intervention
- Any psychological intervention versus no intervention or care as usual

4. Clinical interventions, psychological plus antidepressant medication

- Psychological intervention combined with antidepressant medication versus antidepressant medication alone
- Psychological intervention combined with antidepressant medication versus no intervention or care as usual

5. Clinical interventions, exercise

- Exercise intervention versus any other exercise intervention
- Exercise intervention versus no intervention or care as usual

Types of outcome measures

In this review, we operationalised reduction in work disability as a reduction in sickness absence and as enhancement in work functioning.

Primary outcomes

The main outcome measure in this review was days of sickness absence during the follow-up period. Sickness absence data could be extracted from the employee attendance records or the files of a compensation board, or could be self-reported.

Secondary outcomes

When available, we included the following secondary outcomes from the included studies.

1. Depression (either dichotomously or continuously measured).
2. Work functioning (Nieuwenhuijsen 2010). Examples of work functioning measures are the Endicott Work Productivity Scale (EWPS) (Endicott 1997) and the Sheehan Disability Scale (SDS) (Sheehan 1996). We only included instruments that separately measured work functioning (instead of work and other activities combined).
3. Employment status after a period of time (categories being: 'not working', 'working less hours than the contract hours or having modified duties', or 'working all contract hours without modified duties'.

We did not include other outcomes such as employee satisfaction, general social functioning (not work specific), or quality of life scales.

We considered the effects measured with all the above instruments on the following timescales:

- short term, up to one month;
- medium term, from one month to a year; and
- long term, over a year.

Search methods for identification of studies

Electronic searches

The updated search included all publications from January 2006 up until January 2014 (Appendix 1). For this update, we searched the following electronic databases: CENTRAL (The Cochrane Library), MEDLINE, PsycINFO, EMBASE, and CINAHL. We used three types of terms: depression-related words combined with work-related words and database-specific methodological filter terms. We adapted search terms for PsycINFO, EMBASE, and CINAHL from the MEDLINE search to fit the specific requirements of those databases. For CENTRAL, we replaced the methodological filter by a filter to identify trials.

We based the selected work-related search terms on previous studies. Work* and occupation* are sensitive single terms used to locate occupational health studies, as advocated by Verbeek (Verbeek 2005). Furthermore, we selected database-specific terms relevant to our objective from a study testing which work-related search terms are best suited for literature searching on chronic disease (rheumatoid arthritis, diabetes mellitus, hearing problems, and depression) and work (Haafkens 2006).

We conducted the original search strategy for the first version of this review in 2006, using no limits on publication date (Appendix 2).

Searching other resources

We checked the reference lists of all articles that we retrieved as full papers and of all retrieved systematic and narrative reviews in order to identify further potentially eligible studies.

Data collection and analysis

Selection of studies

Pairs of authors (KN, BF, CF, UB, AV) independently reviewed all studies retrieved from the searches for eligibility. If the title and abstract provided sufficient information to decide that the study did not fulfil the criteria for selection, we excluded the study at that point. We excluded studies in this phase only if the study did not include participants with depressive disorders or it was not a controlled intervention study. When it was not clear whether sickness absence was measured, we retrieved the full article before deciding

upon exclusion. We then examined the full text publications of the remaining studies in order to decide which studies fulfilled all inclusion criteria. We documented the reasons for exclusion at that stage. The two authors discussed any disagreement about the inclusion of studies until they reached consensus. If they could not resolve their difference of opinion, they consulted a third author (JV). We had all articles published in languages other than English translated or assessed for inclusion by a native speaker.

Data extraction and management

We constructed a data extraction form that enabled the authors to extract the data from the included studies. For each study, one author filled out the forms and this form was checked by a second author (AN, AV, BF, CF, HH, KN, and UB participated in data extraction) and they solved differences of opinion by discussion. When only a proportion of the study population was workers, we extracted the data for that subgroup from the article. In the case where these data were not reported, we asked the original authors to provide the data for this subgroup. We used the same procedure for studies where only a proportion of the study population was depressed.

Assessment of risk of bias in included studies

Pairs of authors (AN, AV, CH, HH, or UB with either KN or BF) independently assessed the risk of bias of the included studies. We used the following items to assess risk of bias in the included studies: random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), and selective reporting (reporting bias). We evaluated risk associated with incomplete outcome data or with blinding of outcome assessments separately for depressive symptoms and the sickness absence data. We assessed the risk of bias in RCTs and cluster RCTs by using the Cochrane Collaboration's risk of bias tool (Higgins 2011).

With regard to the risk of attrition bias, we calculated the percentage lost to follow up taking the number randomised as the starting point and the number analysed at the latest follow-up measurement as the endpoint. We assigned a high risk of attrition bias to studies with a percentage of participants lost to follow up of more than 20%, and a low risk for studies with less than 10% lost to follow. The risk of attrition bias for studies with 10% to 20% lost to follow up depended on whether the analyses of results accounted for attrition sufficiently.

We rated each potential source of bias as 'high risk' of bias, 'low risk' of bias, or 'unclear risk' of bias in the 'Risk of bias' table. Next, we constructed a 'Risk of bias' summary figure together with an overview 'risk of bias' graph as described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). Where information on risk of bias related to unpublished data or correspondence with a researcher, we noted this in the risk of bias table.

Measures of treatment effect

We plotted the results of each trial as means and standard deviations (SD) for continuous outcomes. For each timescale (short term, medium term, and long term), we selected the last available observation within this period for the meta-analysis. For the primary outcome measure, that is days of sickness absence, we transformed the number of days or hours worked during the follow up into days of sickness absence. To do so, we extracted the hours or days worked from the maximum of hours an employee would work in that specific country. When transforming the data from days worked to days not worked, the SDs did not need to be transformed. When transforming the data from hours to days, we divided both the means and SDs by eight. Studies used different time spans during which they measured the number of days of sickness absence. Therefore, for days of sickness absence we used the standardised mean difference (SMD) with a 95% confidence interval (CI)) between the intervention and control groups as the summary effect measure.

For the secondary outcome measures, we also used SMDs because it is likely that these outcomes were measured with different instruments. We chose to treat ordinal variables using a scale of more than five categories as continuous variables (it should be noted that this choice was based on arbitrary criteria). We dichotomised scales with less than five categories. For dichotomous data, we calculated the risk ratios (RRs) and 95% CIs.

For depression data, where studies presented both dichotomous and continuous data, we preferred the continuous outcome measures since the majority of the studies presented these.

Unit of analysis issues

For studies that employed a cluster randomised design and did not consider the design effect in the analyses, we planned to calculate the design effect by following the methods stated in the *Cochrane Handbook for Systematic Reviews of Interventions* (Donner 2002) based on a fairly large assumed intra-cluster correlation of 0.10. However, the cluster RCTs included in the review reported negligible intra-cluster correlations. Therefore, we did not adjust the measures of effect presented by the authors.

Dealing with missing data

If the SDs (continuous data) or numbers of outcomes for each group (dichotomous data) were not presented in the publication, we contacted the authors with a request to provide these data. Whenever authors were unable or unwilling to provide this information, we calculated SDs from P values and CIs following the instructions of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).

We sought additional information regarding study details or statistical data, or both, from the authors of 20 studies and received information from 15 authors. Ten of the authors provided statistical data that had not been published in their articles, which enabled us to include nine of these studies in the meta-analyses. In the case of two studies the correspondence led to the exclusion of the study because essential information on the primary outcome measure could not be provided (Simon 2000; Stant 2009). Whenever essential information concerning the risk of bias could not be obtained within four weeks of contacting the authors, we listed the corresponding details as 'unclear'.

Assessment of heterogeneity

We assessed statistical heterogeneity in the meta-analyses with the I^2 statistic. If we observed considerable heterogeneity (I^2 > 75%), we refrained from statistical pooling of the studies within that comparison. Substantial inconsistency (I^2 statistic) also led to downgrading of the quality of the evidence (see Data synthesis for details).

Assessment of reporting biases

We planned to produce funnel plots for visual inspection of possible publication bias. However, due to the small number of studies in each comparison we did not perform these.

Data synthesis

For each predefined comparison, we analysed data for each outcome measure separately. Whenever interventions belonged to the same category in the comparison but two authors (KN and JV, or KN and BF) judged them to be dissimilar, we defined subcategories for these types of intervention. We conducted meta-analysis if two authors (KN and BF) judged a group of trials sufficiently homogeneous in terms of participants, interventions, and outcomes to provide a meaningful summary. In such cases we calculated pooled SMDs for the predefined outcome measures using the Review Manager software (RevMan 2012) with a random-effects model. For three-armed trials contributing evidence to two different comparisons, we divided the number of participants of the arm used in both comparisons by two.

We used the GRADE approach to assess the quality of a body of evidence regarding the primary outcome category of the comparisons addressed in the review. At the start of the GRADE assessment process we assumed high quality for all studies and we downgraded the quality of the evidence for each comparison by one to three levels depending on the seriousness of the violations in each domain.

To assess the risk of bias for a comparison, we considered the risk of bias tables for each study in that comparison. We saw Items related to selection bias, detection bias, and attrition bias as prerequisites for high quality. We only considered studies with low risks on these items to have a low risk of bias. For each comparison we considered the risk of bias serious (-1) if a majority of the evidence in the studies included in the meta-analysis (in terms of weights) were of low quality. We applied a -2 downgrade in cases where the majority of the studies did not have adequate random sequence generation and allocation concealment. For consistency, we considered an I^2 value of 50% to 75% to indicate substantial inconsistency, which lead us to downgrade (-1). If the I^2 value exceeded 75%, we refrained from pooling the results and we analysed the results for each study separately. Indirectness of the evidence was not an issue in our review as all comparisons in the included studies directly addressed the comparison. For imprecision of results, we judged serious imprecision leading to downgrading (-1) if a comparison either included a number of fewer than 400 participants or a wide CI around the effect estimate. For a non-significant effect, we considered a CI to be wide if it included an SMD of both 0 and a moderate effect size (SMD > 0.5 or < -0.5). For a significant effect, we considered a CI to be wide if it included both a small and large effect size (SMD small = -0.2 or 0.2; SMD large = 0.8 or -0.8). We could not detect publication bias in our review due to the low number of studies per comparison.

The resulting interpretation of the quality of the level of evidence per comparison was as follows.

High: further research is very unlikely to change our confidence in the estimate of effect.

Moderate: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low: any estimate of effect is very uncertain.

We created a 'Summary of findings' table with GRADEpro software (GRADEpro 2008) for the main comparisons using the primary outcome categories.

Sensitivity analysis

We planned to conduct sensitivity analyses by excluding:

1. low quality studies,
2. studies with skewed data,
3. cluster randomised trials, and
4. studies in which workers were a small subgroup of the study population.

However, the small numbers of studies in each comparison did not allow for this.

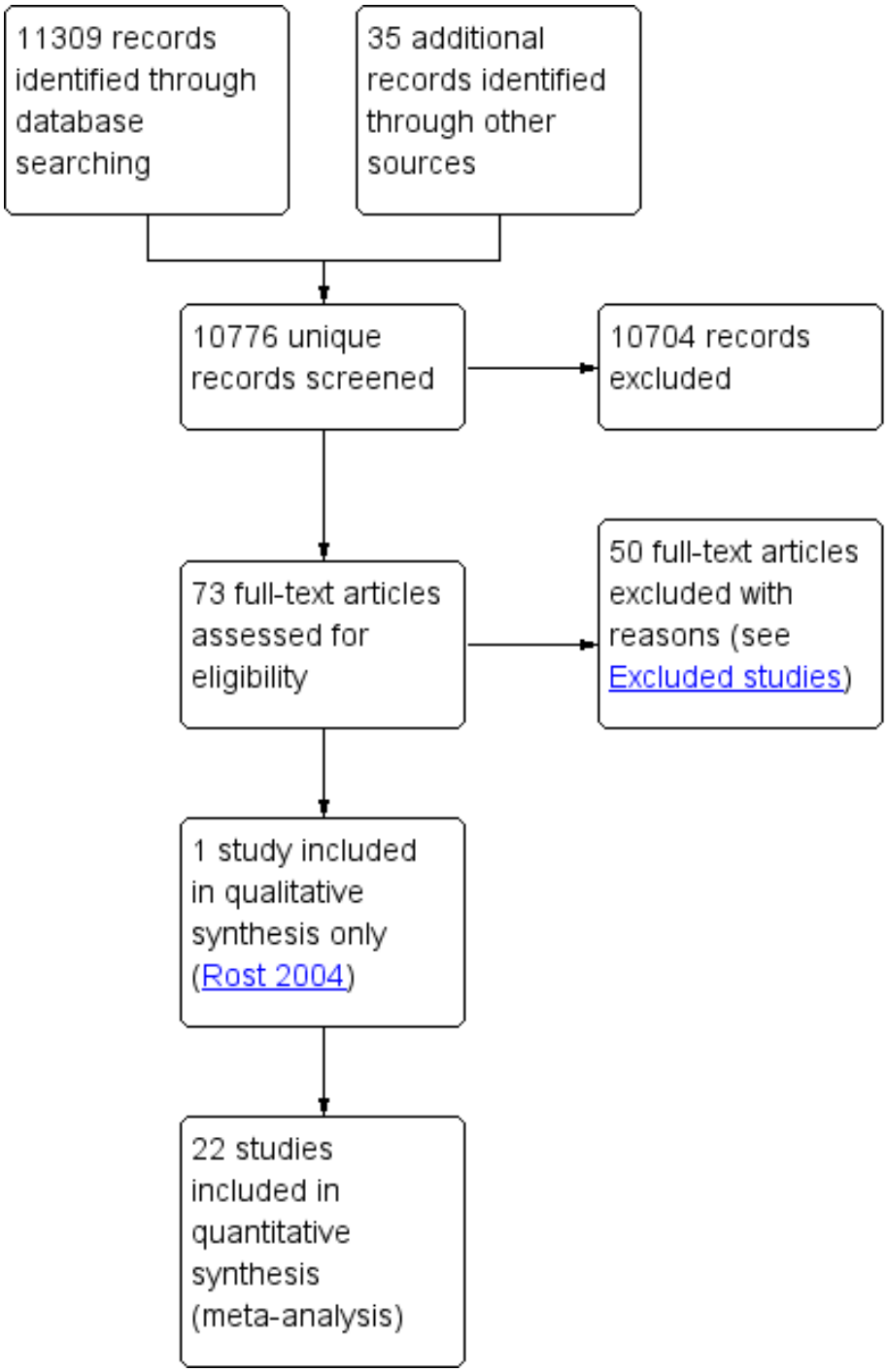
RESULTS

Description of studies

Results of the search

Figure 1 displays a PRISMA study flow chart of the inclusion process. The original and updated electronic searches resulted in 6392 plus 4917 hits. We assessed the titles and abstracts of these combined searches (n = 11,776) for eligibility. This resulted in the full text assessment of 73 (30 plus 43) publications. We excluded fifty (19 plus 31) studies after further scrutiny (see Characteristics of excluded studies). In addition, we identified five ongoing studies (see Characteristics of ongoing studies).

Figure 1. PRISMA Study flow diagram of the study selection process.



Included studies

We included 23 studies in the review (see Characteristics of included studies). Three of these studies included three study arms (Kendrick 2005; Knekt 2013; Krogh 2009). Therefore, we included a total of 26 intervention groups in this review.

Designs

Of the included studies, 20 were RCTs and three were cluster RCTs (Noordik 2013; Rost 2004; Schoenbaum 2001). Intra-class correlations for these studies were reported to be negligible and therefore we did not adjust the data.

Sample sizes

The total number of participants in the included studies was 6278. The number of participants included in the analysis was lower (5996) as we reported on the subgroup of 'employed and depressed participants only' in cases where studies included other subgroups as well. The number of participants in the smallest intervention (sub)group was lower than 20 in one study, between 20 and 100 in 13 studies, between 100 and 200 in six studies, and more than 200 in three.

Time period, setting and participants

Three studies were published before 2000, seven between 2000 and 2005, and 13 after 2005. Five studies were conducted in the US, while 18 were conducted in Europe. Participants were recruited in primary care settings (seven studies), outpatient settings (10 studies), workplace settings (two studies), occupational health care (two studies), a managed care setting (one study), and one study was conducted in a community mental health centre. In 18 studies, all participants had a major depressive disorder. In five studies (Bee 2010; Kendrick 2005; Knekt 2013; McCrone 2004; Noordik 2013) depressed patients constituted a subgroup of the study participants.

Interventions

Work-directed interventions

We identified five work-directed interventions (Hees 2013; Lerner 2012; Noordik 2013; Schene 2006; Vlasveld 2013). The first three compared the addition of a work-directed intervention to clinical care with clinical care alone, whereas one study (Vlasveld 2013) compared work-directed and clinical care with work-directed care alone, and another study (Noordik 2013) compared two alternative work-directed interventions. The Hees 2013 and Schene 2006 studies compared fairly similar interventions, an occupational therapy program focusing on work reintegration and combining modified work and supportive interventions. The intervention in both studies included contact with the occupational physician and the employer, exploration and solving of work problems, and preparation and start of work reintegration. The Lerner 2012 study compared an extension of Employee Assistance Counselling (EAP) to regular EAP. This program incorporated both work modification and support and consisted of: 1) work coaching and modification, 2) care co-ordination, and 3) cognitive behavioural strategies. The Vlasveld 2013 study compared the addition of enhanced clinical care to regular support by the occupational physician who had been trained and had access to the psychiatric consultation. The program contained the following elements: 6 to 12 sessions of Problem Solving Therapy, manual-guided self-help, a workplace intervention and, depending on patient preference, prescription of antidepressant medication according to a treatment algorithm. The Noordik 2013 study compared an exposure-based return to work intervention (RTW-E) conducted by occupational physicians (OPs), gradually exposing the participants to more demanding work situations, to regular support by the OP. The RTW-E program provided workers with several homework assignments aimed at preparing, executing, and evaluating an exposure-based RTW plan. The work-directed 'care as usual' by OPs in the Vlasveld 2013 and Noordik 2013 studies was based on a national guideline and usually included both work modification and support.

Clinical interventions

Antidepressant medication

Six studies examined the effectiveness of antidepressant medication. Three studies compared a SSRI with SNRI medication (Fernandez 2005; Romeo 2004; Wade 2008), one study compared a SSRI with TCA (Miller 1998), one study compared two different SSRIs (Fantino 2007), and a fourth study compared TCA or MAO inhibitors with placebo (Agosti 1991).

Psychological interventions

One study (Knekt 2013) with three study arms compared two psychological interventions (short-term and long-term psychodynamic psychotherapy) with an alternative psychological interven-

tion, solution-focused therapy. Four studies looked at the effects of a specific psychological intervention as compared to care as usual (Bee 2010; Hollinghurst 2010; Kendrick 2005; McCrone 2004). The Kendrick 2005 study had three intervention arms and compared two types of psychological interventions performed by community mental health nurses (Problem Solving Therapy and generic counselling) with general practitioner care as usual. Two studies (Hollinghurst 2010; McCrone 2004) compared a computerised form of CBT with general practitioner care as usual. In the Hollinghurst 2010 study, participants receiving online CBT were offered up to 10 sessions each having a duration of 55 minutes. Each participant was assigned their own therapist for the duration of the study. Participants and therapists typed free text into the computer, with messages sent instantaneously, and only used this means of communication. Online CBT in the McCrone 2004 study included a 15-minute introductory video, eight 50-minute sessions of CBT, and homework projects between sessions. The program was interactive and feedback was provided to both the patient and general practitioner after each session. One study (Bee 2010) compared telephone CBT to usual primary and occupational health services. Telephone CBT was delivered over 12 weeks by registered graduate mental health workers. Participants worked with therapists through regular phone calls to identify and challenge negative thoughts, develop self-care skills, and complete workbook exercises emphasizing behavioural activation.

Psychological interventions plus antidepressant medication

Five studies included interventions with a combination of psychological interventions and antidepressant medication. One study (Burnand 2002) compared the effect of psychodynamic therapy combined with TCA medication with TCA medication alone. The intervention included individual sessions by a nurse combined with clomipramine for a duration of 10 weeks. The frequency of the psychotherapy sessions was not fixed. This was compared to a group receiving the same medication and who received supportive care (an individual session with empathic listening, guidance, and support).

Three studies (Rost 2004; Schoenbaum 2001; Simon 1998) compared enhanced primary care with primary care as usual. In these types of interventions general practitioners were enrolled in a quality improvement program and were expected to provide enhanced care including antidepressant medication and psychological interventions, according to primary care guidelines.

One study (Wang 2007) compared a structured telephone outreach and care management program to usual managed care. The telephone outreach systematically assessed needs for treatment, facilitated entry into in-person treatment (both psychotherapy and antidepressant medication), monitored and supported treatment adherence, and (for those declining in-person treatment) provided a structured psychotherapy intervention by telephone. Intervention participants declining in-person treatment and experiencing significant depressive symptoms after two months were offered a structured eight-session cognitive behavioural psychotherapy program.

Exercise

Two studies (Krogh 2009; Krogh 2012), of which the first one included three study arms, looked at the effect of exercise interventions. Krogh 2009 compared supervised strength training or aerobic training to relaxation training. The strength training was designed to increase muscular strength. The training was a circuit training program with six exercises involving large muscle groups on machines. The aerobic training was designed to increase fitness, as measured by maximal oxygen uptake. The program involved 10 different aerobic exercises using large muscle groups. Machines were used for cycling, running, stepping, abdominal exercises, and rowing. In both exercise interventions, all patients were scheduled to meet twice per week during a four-month period for a total of 32 sessions. The relaxation training was designed to avoid muscular contractions or stimulation of the cardiovascular system and included exercises on mattresses followed by light balance exercises and by relaxation exercises with alternating muscle contraction and relaxation in different muscle groups while lying down.

In the Krogh 2012 study, aerobic training was compared to an attention control group (stretching exercises at low intensity). The aerobic training was designed to increase fitness as measured by maximal oxygen uptake. After an initial 10 minutes of general low-intensity warm-up, the participants did 30 minutes of aerobic exercise on a stationary cycle ergometer followed by a five-minute low-intensity cool down period. Both groups were scheduled to meet three times per week for three months, for a total of 36 sessions.

Outcomes

Studies were only selected if they reported on sickness absence. Of the 23 included studies, six studies (Agosti 1991; Bee 2010; Krogh 2012; Miller 1998; Schene 2006; Wang 2007) reported days or hours worked instead of days of sickness absence. These measures were transformed into days of sickness absence as described in the 'Methods' section (see Measures of treatment effect).

We were able to collect data on depression for all but one of the included studies (Agosti 1991). Of all studies reporting on depression, one study (Schoenbaum 2001) presented only dichotomous depression data while all others presented continuous data.

Eight studies (Agosti 1991; Burnand 2002; Hees 2013; Lerner 2012; Miller 1998; Rost 2004; Wade 2008; Wang 2007) reported on work functioning using a (sub)scale that separately measured work instead of work and other activities combined.

None of the included studies reported on employment status after a period of time with the predefined categories: 'not working', 'working fewer hours than the contract hours or having modified

duties' or 'working all contract hours without modified duties'. However, three studies (Krogh 2012; Schoenbaum 2001; Wade 2008) did report 'not working' or 'working' at the end of follow up.

Follow up

(a) Short term

None of the included studies had the last outcome measurement within one month.

(b) Medium term

In 19 studies the last follow-up measurement was between one month and a year after inclusion. Four studies had the last follow-up measurement later than one year but provided data on earlier time points as well (Hees 2013; Knekt 2013; Rost 2004; Schene 2006). We included these outcomes in the medium-term analysis. We used the last available observation within the first year for this purpose.

(c) Long term

In five studies, the last follow-up measurement was later than one year after inclusion. One study reported on a follow-up period of 18 months (Hees 2013), two on 24 months (Rost 2004; Schoenbaum 2001), one on 42 months (Schene 2006), and one on five years (Knekt 2013). However, only depression data and not the days of sickness absence were reported at two years in the Schoenbaum study. We therefore refrained from using the depression data at this time point, leaving four studies with long-term outcome data.

Excluded studies

We excluded a total of 50 studies from the review. Reasons for excluding studies were:

- sickness absence not measured as an outcome (Ahola 2012; Amore 2001; Barbui 2009; Boyer 1998; Brandes 2011; Carlin 2010; Castillo-Pérez 2010; Dunlop 2011; Erkkilä 2011; Finley 2003; Hirani 2010; Kojima 2010; Kroenke 2001; Kuhs 1996; Lam 2012; Martinez 2011; Meyer 2009; Mundt 2001; Oakes 2012; Salminen 2008; Sandahl 2011; Simon 2000; Sir 2005; Stant 2009);
- participants had a mild depressive disorder or were not diagnosed with a depressive disorder at all (Aelfers 2013; Bakker 2007; Blonk 2007; Brouwers 2007; Furukawa 2012; Hackett 1987; Lagerveld 2012; Lexis 2011; Mino 2006; Morgan 2011; Zeeuw 2010);
- not a RCT design (Bech 2000; Eklund 2012; Knekt 2011; Schmitt 2008; Zambori 2002);
- no worker population (Alexopoulos; Folke 2012; Forman 2012);
- study took place in an inpatient care setting (Dick 1985; Hordern 1964);
- participants had a severe mental disorder such as schizophrenia (Becker 1998);
- not able to define a subgroup of depressed patients (Gournay 1995); and
- a double publication (Schoenbaum 2002; Wells 2000).

Studies awaiting assessment

There were five ongoing studies awaiting further assessment (Beurden 2013; Geraedts 2013; Heer 2013; Hellstrom 2013; Warmerdam 2007).

Risk of bias in included studies

We judged studies to have an overall high risk of bias when the items for random sequence generation, allocation concealment, and incomplete outcome data for our primary outcome measure all scored a rating of high risk of bias. We considered the overall risk of bias to be high in 14 studies (Agosti 1991; Burnand 2002; Hollinghurst 2010; Kendrick 2005; Knekt 2013; Krogh 2009; Miller 1998; Noordik 2013; Romeo 2004; Rost 2004; Schene 2006; Schoenbaum 2001; Simon 1998; Wade 2008), Of these 14 studies, six studies (Agosti 1991; Burnand 2002; Miller 1998; Noordik 2013; Rost 2004; Schoenbaum 2001) had either unclear or inadequate random sequence generation or allocation concealment, causing us to classify these studies as having a very high risk of bias. We considered nine studies to have an overall low risk of bias (Bee 2010; Fantino 2007; Fernandez 2005; Hees 2013; Krogh 2012; Lerner 2012; McCrone 2004; Vlasveld 2013; Wang 2007). See Figure 2 and Figure 3 for an overview of the risk of bias per study and the 'Risk of bias' tables that form part of the Characteristics of included studies for details.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

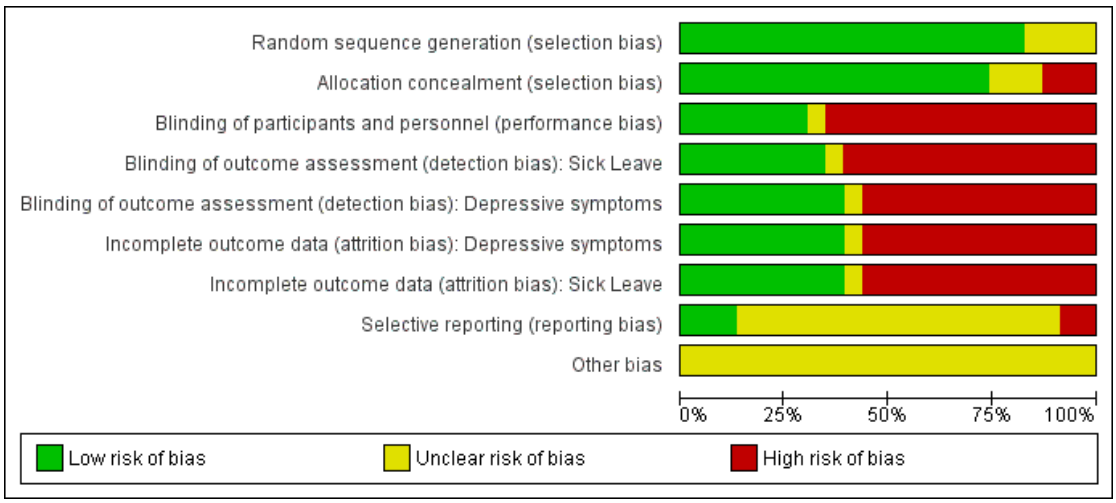
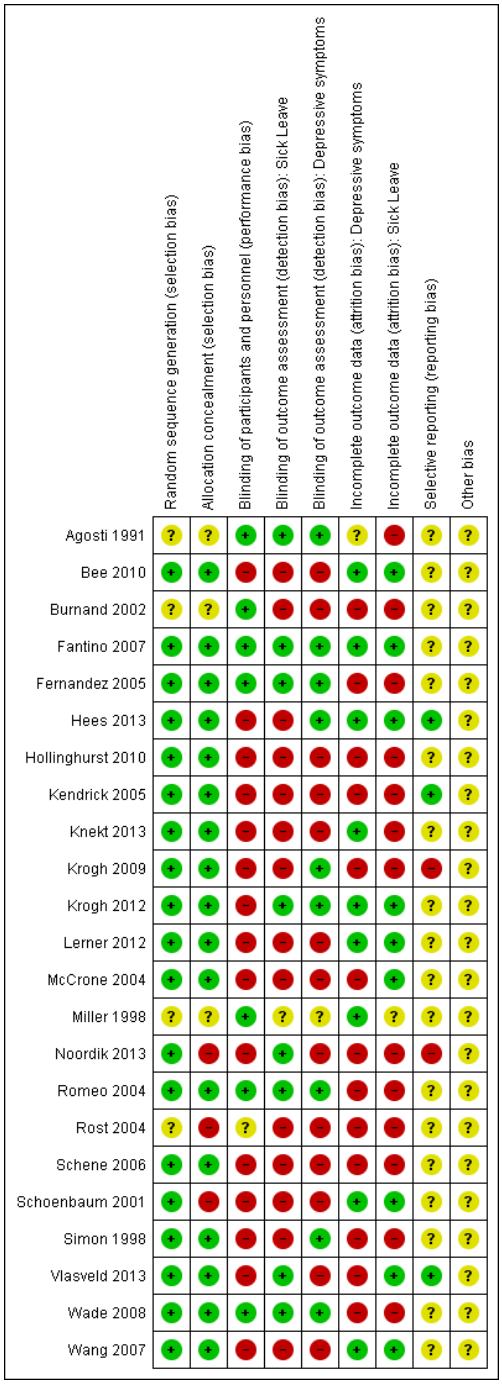


Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.



Allocation

The method for generating random numbers was not adequately described in eight studies. However, personal communications with the authors revealed that in four studies the chosen method was adequate. In the other four studies this information could not be retrieved, which led us to rate the studies as having an unclear risk of bias for this item.

In the three cluster RCTs (Noordik 2013; Rost 2004; Schoenbaum 2001) allocation concealment was not adequate, which was probably indicative of the non-feasibility of allocation concealment in this type of design. In three further studies (Agosti 1991; Burnand 2002; Miller 1998) information on allocation concealment could not be retrieved, leading to a judgment of unclear risk of bias.

Blinding

Risk of performance bias was low in studies using a double-blind design (blinding of participant and care provider). This design was feasible in studies comparing the occupational health effects of antidepressant medications. This type of study has a low risk of performance bias (Agosti 1991; Fantino 2007; Fernandez 2005; Miller 1998; Romeo 2004; Wade 2008). In work-directed, psychological, or exercise interventions blinding of the participant or care provider is not feasible. However, we considered the risk of performance bias high only in those studies where the control intervention could be considered less desirable by participants or care provider. One study (Burnand 2002) managed to compose two evenly desirable psychological interventions, leading to an assessment of low risk of performance bias.

Our primary outcome measure (days of sickness absence) could be measured either by self-report or retrieval from attendance records. In the case of self-report, the outcome could be biased by unblinded participants' knowledge of the intervention. In 15 studies we considered the risk of detection bias to be high, and in one case this risk was unclear.

Incomplete outcome data

We found nine of the 23 studies to have a low risk of attrition bias, with some studies (Knekt 2013; McCrone 2004; Vlasveld 2013) showing different levels of risk of bias for sickness absence and depressive symptoms. Studies with attrition between 10% and 20% could still be classified as having low risk of attrition bias if adequate analyses were conducted to take selective attrition into account. Examples of such analyses are multiple imputation methods or sensitivity analyses.

Selective reporting

For the majority of the studies (19), no design paper or trial registration could be identified in order to assess the risk of selective reporting. In three studies we considered the risk to be low (Hees 2013; Kendrick 2005; Vlasveld 2013) and in one study an outcome measure that was presented in the study design was not reported as an outcome (Noordik 2013).

Other potential sources of bias

We did not identify other potential sources of bias.

Effects of interventions

See: **Summary of findings for the main comparison** Work-directed intervention plus clinical intervention compared to clinical intervention alone for depressive disorder; **Summary of findings 2** Any psychological intervention versus no intervention or care as usual for depressive disorder; **Summary of findings 3** Psychological intervention combined with antidepressant medication versus no intervention or usual care for depressive disorder

The 23 studies included in the review examined work-directed and clinical interventions. The clinical interventions studied were antidepressant medication, psychological or exercise intervention, or a combination of two. We present summary of findings tables for the comparisons with more than two included studies. Table 1 presents the GRADE assessment of the quality of the evidence per comparison.

We did not identify any studies for the comparisons: 'any work-directed intervention versus no intervention or care as usual', 'any antidepressant medication versus any psychological intervention', 'exercise intervention versus any other exercise intervention'. We refrained from conducting sensitivity analyses due to the small number of studies in each comparison.

Below we present the results for our primary outcome, sickness absence, for each of the comparisons. We present our secondary outcomes, depressive symptoms and work functioning, for each of the work-directed interventions as well. For the clinical interventions we chose to only present the effect on depressive symptoms if the intervention reduced sickness absence.

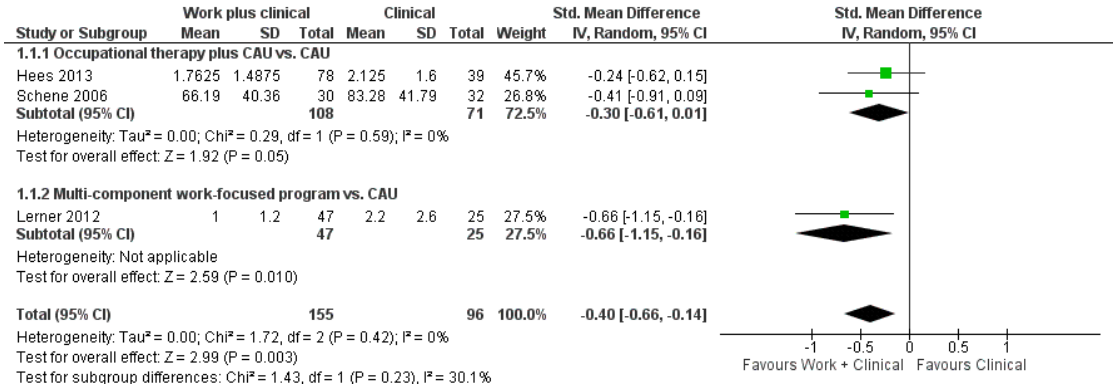
1. Work-directed interventions

1.1 Work-directed intervention combined with a clinical intervention versus clinical intervention alone (medium term)

Three studies looked at the effectiveness of a work-directed intervention combined with a clinical intervention in comparison

to a clinical intervention alone (Hees 2013; Lerner 2012; Schene 2006). The summarised sickness absence results showed moderate quality evidence of a positive effect of adding a work-directed intervention to a clinical intervention (SMD -0.40; 95% CI -0.66 to - 0.14). The two studies adding occupational therapy to clinical depression care (Hees 2013; Schene 2006) alone did not find a statistically significant effect on sickness absence (SMD -0.30; 95% CI -0.61 to 0.01), while the single study evaluating a multicomponent work-focused intervention (Lerner 2012) did find a reduction of sickness absence days (SMD -0.66; 95% CI -1.15 to -0.16; Analysis 1.1). See Summary of findings for the main comparison and Figure 4.

Figure 4. Forest plot of comparison: 1 Work-directed plus clinical versus clinical alone (medium term), outcome: 1.1 Days of sickness absence.



The combined results of these three studies showed no difference between the interventions when evaluating depressive symptoms (SMD -0.32; 95% CI -0.88 to 0.25) or work functioning (SMD -0.31; 95% CI -0.79 to 0.16; Analysis 1.2; Analysis 1.3).

1.2 Work-directed intervention combined with clinical intervention versus clinical intervention alone (long term)

Two studies also reported long-term effects. We combined these in a separate comparison (Hees 2013; Schene 2006). These two studies provided moderate quality evidence that adding a work-directed intervention to a clinical intervention did not reduce sickness absence in the long term (SMD -0.19; 95% CI: -0.49 to 0.12; Analysis 2.1). However, one of the two studies (Hees 2013) found that the work-directed intervention reduced depressive symptoms in the long term (SMD -0.63; 95% CI -1.02 to -0.24; Analysis 2.2).

1.3 Work-directed intervention combined with clinical intervention versus work-directed alone (medium term)

We included one study in this comparison (Vlasveld 2013). The study compared the addition of enhanced clinical care (collaborative care model) to regular support by the occupational physician only. This single study provided moderate quality evidence that the intervention was not more effective in reducing sickness absence (SMD -0.14; 95% CI -0.49 to 0.21) or depressive symptoms (SMD 0.26; 95% CI -0.20 to 0.72) than the control intervention (Analysis 3.1; Analysis 3.2).

1.4 Any work-directed intervention versus alternative work-directed intervention

We included one study (Noordik 2013) in this comparison; an exposure-based return to work program was compared to regular occupational physician support. The study provided very low quality evidence that sickness absence could have been reduced more in the control group but this effect was not statistically sig-

nificant (SMD 0.45; 95% CI -0.00 to 0.91; Analysis 4.1). The exposure-based return to work program also did not reduce depressive symptoms (SMD -0.18; 95% CI -0.84 to 0.49; Analysis 4.2).

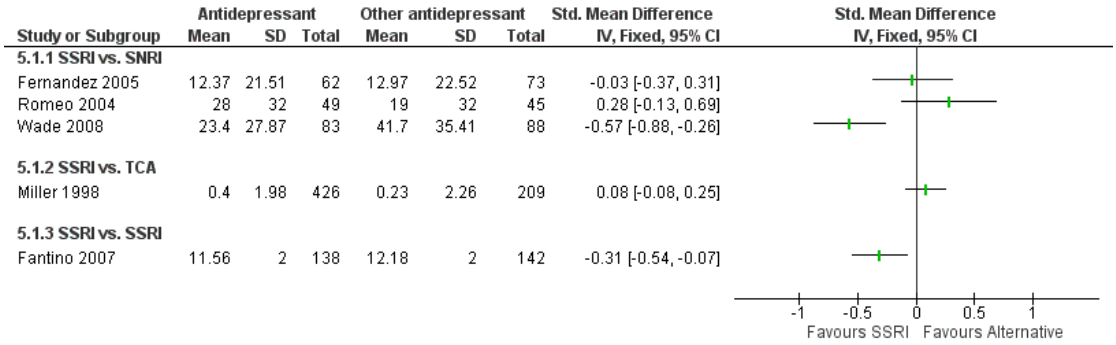
2. Clinical interventions, antidepressant medication

2.1 Any antidepressant medication versus any other antidepressant medication (medium term)

2.1.1 SSRI versus SNRI

Three studies compared a SSRI to SNRI in depressed workers (Fernandez 2005; Romeo 2004; Wade 2008). In the meta-analysis, the inconsistency of results between these three studies (I^2) was 83% and so we did not pool them (Figure 5). The results of the single studies were highly inconsistent. We found no difference in sickness absence between a SSRI and SNRI in the Fernandez 2005 study (SMD -0.03; 95% CI -0.37 to 0.31) as well as in the Romeo 2004 study (SMD 0.28; 95% CI -0.13 to 0.69). The Wade 2008 study revealed evidence of an effect on sickness absence favouring a SSRI (SMD -0.57; 95% CI -0.88 to -0.26; Analysis 5.1). Measured with the Sheehan disability scale, this study also reported a favourable effect on work functioning (difference of 2.4; 95% CI 0.4 to 4.1) but the reported data did not allow for inclusion in the meta-analysis.

Figure 5. Forest plot of comparison: 5 Any antidepressant medication versus any other antidepressant medication, outcome: 5.1 Days of sickness absence.



2.1.2 SSRI versus TCA

Miller 1998 was the only study comparing a SSRI to TCA medication in depressed workers. This study found no difference between a SSRI and TCA in reducing sickness absence (SMD 0.08; 95% CI -0.08 to 0.25; Analysis 5.1). The Miller 1998 study measured work functioning using the SAS work composite (Wells 1989). A higher score on this measure reflects a higher level of impairment. The study reported no significant difference on work functioning between the groups (difference of -0.08; 95% CI -0.24 to 0.09; Analysis 5.3).

2.1.3 SSRI versus SSRI

One study (Fantino 2007) compared one SSRI to another SSRI. This study found evidence of a greater reduction in sickness absence with escitalopram compared to citalopram (SMD -0.31; 95% CI -0.54 to -0.07; Analysis 5.1).

2.2 Any antidepressant medication versus placebo

One study compared a TCA or MAO to placebo (Agosti 1991). We found very low quality evidence, based on one study, that antidepressant medication did not reduce sickness absence. The effect may even have been in favour of the placebo condition (SMD 0.48; 95% CI -0.05 to 1.00) but this was not statistically significant (Analysis 6.1). Measured with the Work functioning subscale of the LIFE interview, Agosti 1991 did find a statistically significant positive effect in favour of antidepressant medication (SMD -0.58; 95% CI -1.11 to -0.05; Analysis 6.2).

3. Clinical interventions, psychological

3.1 Any psychological intervention versus other psychological intervention (medium term)

One study (Knekt 2013) with three treatment arms evaluated the effect of alternative psychological interventions. Two study arms assessed psychodynamic therapy, where one study arm examined short-term and the other long-term therapy. Both were compared to solution focused therapy. The inconsistency (I^2) in this meta-analysis was 97%, therefore we refrained from pooling the results of the two psychodynamic therapy conditions.

We found low quality evidence of short-term psychodynamic therapy not being more effective than solution focused therapy in reducing sick leave (SMD 0.25; 95% CI -0.39 to 0.89) and of solution focused therapy being more effective in reducing sick leave than long-term psychodynamic therapy (SMD 1.16; 95% CI 0.49 to 1.83; Analysis 7.1).

The depressive symptoms and work functioning outcomes were better for the short-term psychodynamic therapy than for solution focused therapy (SMD -0.66; 95% CI -1.03 to -0.30), but solution focused therapy showed better results than the long-term psychodynamic therapy (SMD 1.00; 95% CI 0.63 to 1.36; Analysis 7.1; Analysis 7.2).

3.1 Any psychological intervention versus other psychological intervention (long term)

The Knekt 2013 study also had long-term results (five-year follow up). We refrained from statistically pooling the results due to high inconsistency (I^2 = 99%). The separate analyses yielded low quality evidence of long-term (SMD -0.91; 95% CI -1.62 to -0.19) and short-term psychodynamic psychotherapy (SMD -4.61; 95% CI -5.84 to -3.39) reducing sickness absence more effectively than solution focused therapy in the long term (Analysis 8.1).

3.1 Any psychological intervention versus no intervention or care as usual (medium term)

3.1.1 Online or telephone CBT versus care as usual

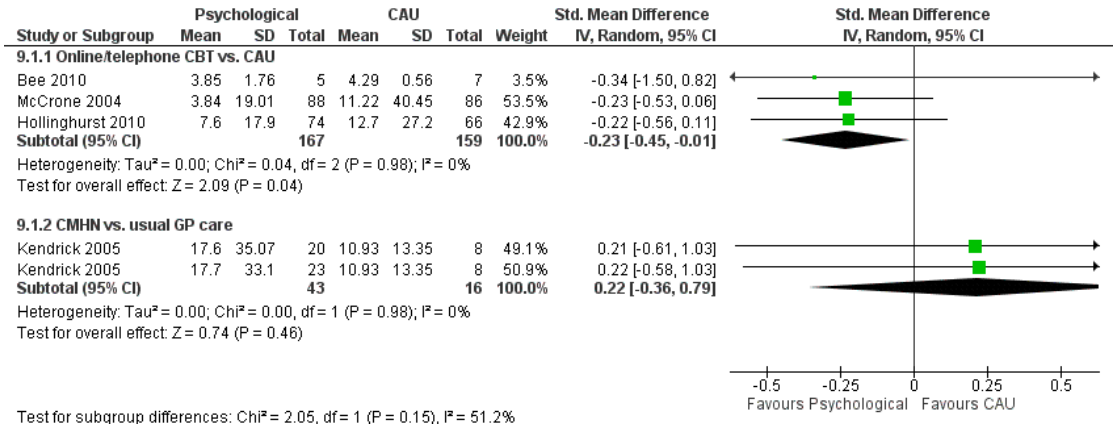
Three studies compared online or telephone CBT with care as usual (Bee 2010; Hollinghurst 2010; McCrone 2004). The pooled results showed that there was moderate quality evidence (SMD -0.23; 95% CI -0.45 to -0.01) that online or telephone CBT reduced sickness absence more than usual primary or occupational care. See Summary of findings 2 and Analysis 9.1. Online or telephone CBT also reduced depressive symptoms (SMD -0.56; 95% CI -0.76 to -0.36; Analysis 9.2).

3.1.2 Problem solving or counselling by community mental health nurse versus general practitioner

One study (Kendrick 2005) looked at two types of psychological interventions performed by community mental health nurses (Problem Solving Therapy and generic counselling) and compared these interventions with general practitioner care as usual. The pooled results from the two psychological interventions yielded low quality evidence (SMD 0.22; 95% CI -0.36 to 0.79) that these interventions were no better in reducing sickness absence than care by general practitioners (Analysis 9.1).

See Figure 6 for the forest plot of this comparison.

Figure 6. Forest plot of comparison: 9 Any psychological versus no intervention or care as usual, outcome: 9.1 Days of sickness absence.



4. Clinical interventions, psychological plus antidepressant medication

4.1 Psychological intervention combined with antidepressant medication versus antidepressant medication alone (medium term)

One study (Burnand 2002) evaluated the effectiveness of psychodynamic therapy combined with TCA medication versus TCA medication alone. This study provided very low quality evidence of psychodynamic therapy combined with TCA medication reducing sickness absence more than TCA medication alone (SMD -0.71; 95% CI -1.25 to -0.17; Analysis 10.1).
The effects of the intervention on work functioning (SMD -0.11; 95% CI -0.57 to 0.35) and depressive symptoms (SMD -0.49; 95% CI -1.02 to 0.04) were both not significant (Analysis 10.2; Analysis 10.3).

4.2 Psychological intervention combined with antidepressant medication versus no intervention or care as usual (medium term)

The findings for this comparison are displayed in Summary of findings 3.
Three studies tested enhanced primary care interventions, which were deemed similar enough for statistical pooling (Rost 2004; Schoenbaum 2001; Simon 1998). However, for the study by Rost 2004 we could not obtain from the author the SDs around the mean estimates, nor could we calculate them. Therefore, we could not include this study in the meta-analysis. Because the publication itself only presented data over two years, we qualitatively described the results of this study for the long-term outcome only.
Both trials with usable data (Schoenbaum 2001; Simon 1998) failed to show a significant difference in days of sickness absence between the intervention and comparison groups. The pooled results of these two trials provided low quality evidence of no effect of enhanced primary care on sickness absence (SMD -0.02; 95% CI -0.15 to 0.12; Analysis 11.1). In addition, Schoenbaum 2001 did not find a significant difference in employment status between the intervention and the control groups in the medium term (RR 1.08; 95% CI 0.99 to 1.18; Analysis 11.2).

One study (Wang 2007) looked at a different type of intervention, a structured telephone outreach and care management program, in comparison to usual care. We found high quality evidence of an effect on sickness absence in favour of the intervention (SMD -0.21; 95% CI -0.37 to -0.05; Analysis 11.1) based on this study. The effect on depressive symptoms was similar (SMD -0.23; 95% CI -0.39 to -0.07; Analysis 11.3). However, the effect on work functioning favoured the control condition (SMD 0.50; 95% CI 0.34 to 0.66; Analysis 11.5).

4.3 Psychological intervention combined with antidepressant medication versus no intervention or care as usual (long term)

Only one study (Rost 2004) reported long-term outcomes. Data were insufficient for us to calculate a SMD in days of sickness absence. The authors reported no statistically significant effect on sickness absence and depression but did report that work functioning was significantly improved using a subjective rating on a 0 to 10 scale at work.

5 Clinical interventions, exercise

5.1 Exercise intervention versus no intervention or care as usual

We included two studies in this comparison (Krogh 2009; Krogh 2012), of which one (Krogh 2009) had two study arms.

5.1.1 Strength exercise versus relaxation

We found low quality evidence, based on one study, that supervised strength exercise was more effective than relaxation in reducing sickness absence (SMD -1.11; 95% CI -1.68 to -0.54; Analysis 12.1).

5.1.2. Aerobic exercise versus relaxation or stretching

The pooled effect of two studies yielded moderate quality evidence that aerobic exercise was not more effective than relaxation or stretching in reducing sickness absence (SMD -0.06; 95% CI -0.36 to 0.24; Analysis 12.1).

ADDITIONAL SUMMARY OF FINDINGS *[Explanation]*

Any psychological intervention versus no intervention or care as usual for depressive disorder

Patient or population: Persons with depressive disorder
Settings: One study was conducted in a workplace setting and two in primary care
Intervention: Any psychological intervention versus no intervention or care as usual (medium term)

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Care As Usual (CAU)	Any psychological intervention				
Days of sickness absence Follow up: 3 - 8 months		The mean days of sickness absence in the intervention groups was **0.23 standard deviations lower** (0.45 to 0.01 lower)	SMD -0.23 (-0.45 to -0.01)	326 (3 studies)	⊕⊕⊕◯ **moderate**[1]	A standard deviation of 0.2 represents a small difference between groups

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CBT: Cognitive-Behavioral Therapy; **CAU**: Care As Usual; **CI:** Confidence interval; **SMD**: Standardised Mean Difference

Intervention description
All three interventions were cognitive-behavioral therapy, one by telephone and two online. Each of the interventions were interactive, with therapists or specialised nurses providing feedback
GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Downgraded one level because N < 400

Psychological intervention combined with antidepressant medication versus no intervention or usual care for depressive disorder

Patient or population: Persons with depressive disorders
Settings: Two studies were conducted in a primary care and one in a managed care setting
Intervention: Psychological intervention combined with antidepressant medication versus no intervention or usual care (medium term)

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Care As Usual (CAU)	Psychological intervention combined with antidepressant medication				
Days of sickness absence - Enhanced primary care versus CAU Follow up: 7 - 12 months		The mean days of sickness absence in the intervention groups was **0.02 standard deviations lower** (0.15 lower to 0.12 higher)	SMD -0.02 (-0.15 to 0.12)	969 (2 studies)	⊕⊕◯◯ **low**[1,2]	A standard deviation of 0.2 represents a small difference between groups
Days of sickness absence - Telephone outreach and care management program versus CAU Follow up: mean 12 months		The mean days of sickness absence in the intervention groups was **0.21 standard deviations lower** (0.37 to 0.05 lower)	SMD -0.21 (-0.37 to -0.05)	604 (1 study)	⊕⊕⊕⊕ **high**	A standard deviation of 0.2 represents a small difference between groups

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CAU: Care As Usual; **CI:** Confidence interval; **SMD**: Standardised Mean Difference

Intervention description

Enhanced primary care

General practitioners were enrolled in a quality improvement program and were expected to provide enhanced care including antidepressant medication and psychological interventions, according to primary care guidelines

Telephone outreach and care management

This program systematically assessed needs for treatment, facilitated entry into in-person treatment (both psychotherapy and antidepressant medication), monitored and supported treatment adherence, and (for those declining in-person treatment) provided a structured psychotherapy intervention by telephone

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

[1] Downgraded with one level because all studies were of low quality

[2] Downgraded with one level because in one study allocation concealment not adequate

DISCUSSION

Summary of main results

We included 20 RCTs and three cluster RCTs in the review, with five studies evaluating work-directed interventions and 18 evaluating clinical interventions. Within these broad categories, the type of intervention varied widely from one study to another, which limited the number of studies in each predefined comparison. This review showed that there is moderate quality evidence that adding a work-directed intervention to a clinical intervention reduces the number of days on sick leave in the medium term (4 to 12 months; SMD -0.40). Similar effects on depressive symptoms and work functioning could not be demonstrated. The absolute difference in days absent from work differed in each of the studies. The Hees 2013 study found the smallest absolute difference, 1.7 versus 2.1 days over a 12-month period. Schene 2006 found a difference of 66 versus 83 days absence over a period of six months; and Lerner 2012 found an absolute difference of one versus 2.2 days over a period of two weeks. Moderate quality evidence, from a comparison including a single study, showed that enhancing the clinical care in addition to regular work-directed care did not reduce sickness absence more than work-directed care alone (with regular access to clinical care). There was very low quality evidence, based on one study, that regular care by occupational physicians that is enhanced with an exposure-based return to work program does not reduce sickness absence compared to regular care by occupational physicians.

With regard to antidepressant medication, this review found highly inconsistent results regarding the effect of SSRIs compared to other medications on days of sickness absence (four studies). Compared to SNRI medication (three studies), one single study found that SSRI reduced sickness absence (Wade 2008), no difference in effect on sickness absence was found in another (Fernandez 2005), and a non-significant difference in effect on sickness absence was found in the last (Romeo 2004). One single study found that a SSRI did not reduce sickness absence more than TCA medication (Miller 1998). One study (Fantino 2007) compared one SSRI to another SSRI. This study found that escitalopram reduced sickness absence more than citalopram (SMD -0.31). One study compared a TCA or MAO to placebo (Agosti 1991). This study found that the antidepressant medication did not reduce sickness absence more than placebo.

This review found moderate quality evidence based on three studies that telephone or online CBT reduced sick leave more than usual primary or occupational care in the medium term (three to eight months; SMD -0.23). These interventions also reduced depressive symptoms compared to usual primary or occupational care (SMD -0.56). The absolute differences in days of sickness absence were 3.9 versus 4.3 over four weeks (Bee 2010); 3.8 versus 11.2 over eight months (McCrone 2004); and 7.6 versus 12.7 over eight months (Hollinghurst 2010).

This review found low quality evidence of no effect of enhanced primary care on sickness absence in the medium term (four to 12 months), based on the pooled results of two studies (SMD -0.02). A third study found no statistically significant effect on sickness absence of this intervention in the long term (24 months).

This review found high quality evidence from one study that a structured telephone outreach and care management program reduced sickness absence when compared to usual care (SMD -0.21). The absolute difference on the original scale was 42.3 (intervention) versus 39.5 (control) hours worked per week over the last four weeks.

Overall completeness and applicability of evidence

The studies included in this review have been conducted in Europe and the United States of America only. Therefore, the generalisability of our findings to other parts of the world remains unclear. In line with our inclusion criteria, the included studies cover a range of clinical states. In 13 studies a major depressive disorder according to the DSM-IV or III was used as an inclusion criterion, while others included patients based on their symptom severity as measured by a questionnaire. Moreover, study setting is likely to be a source of clinical heterogeneity. Most studies were conducted in primary care settings (seven studies) and in outpatient settings (10 studies). In only two studies patients were recruited in a workplace setting, and in another two in occupational health care. In many instances the occupation of the participants was not reported even though it is conceivable that the effect of interventions partly depends on the specific work situation. A lack of studies on work-related factors which may be predictors for work outcomes in depressed workers has already been pointed out (Lagerveld 2010). Therefore, we cannot assess the potential impact of work situations on the effectiveness of the included interventions.

In this updated review, we were able to include studies on work-directed interventions as well as clinical interventions. While it is important to assess the effects of clinical interventions on occupational health, we are aware that the primary reason to choose between one or another clinical intervention is clinical effectiveness. However, in line with the emerging paradigm of value-based medicine, it is central to care to offer interventions to patients providing the greatest patient value (Brown 2013). As being able to work may be one of the factors on which patient preference is based, assessing occupational health outcomes for clinical interventions is key. Moreover, from the point of view of patient preference, work functioning may be as important as sickness absence. However, in most included studies this outcome was not measured. Evaluating the effect of interventions on work functioning would further enable us to assess the patient value of these interventions.

In contrast to the first version of this review, we were able to include studies in most of the predefined comparisons. However, the

number of studies within each comparison was small, and even within some of the comparisons the interventions were too dissimilar to pool the results. One example is the comparison 'psychological intervention combined with antidepressant medication versus no intervention or usual care'. Two studies were on interventions indirectly targeting the worker by enhancing the care of the care provider (Schoenbaum 2001; Simon 1998). One other study (Wang 2007) evaluated an intervention in which the workers were directly targeted as they received psychotherapy by telephone. Only the latter intervention reduced the days of sickness absence more than care as usual. Another consequence of the low number of studies per comparison is that we were unable to perform subgroup analyses for participant and intervention characteristics, which impedes generalisation of the results.

The clinical relevance of the observed effects can best be evaluated by looking at the absolute differences in days of sickness absence. It should, however, be noted that these differences vary from one study to another. Part of the explanation is that the outcome measure 'days of sickness absence' is by definition partly determined by the length of follow up. Nonetheless, variations in the absolute difference between studies are not always explained by differences in length of follow up (see Hees 2013; Schene 2006). The relevance of reductions in days of sick leave depends on the perspective of the stakeholder. A reduction in sick leave of one day may not be relevant from the worker's point of view but can be relevant for stakeholders who bear the costs of the lost productivity, such as employers or insurance companies.

Quality of the evidence

Of the included studies, 20 were RCTs and three were cluster RCTs. The number of participants per study was fairly small, less than 100 in 14 studies, between 100 and 200 in six, and more than 200 in three. In some cases the low number of participants was due to our need to focus only on a subgroup of the study population, either disregarding participants with other mental disorders or participants who did not work.

We considered the overall risk of bias to be low in nine studies. In 14 studies we considered it to be high as either random sequence generation, allocation concealment, or incomplete sickness absence data were inadequate. In six of these the random sequence generation or allocation concealment was unclear or inadequate.

In the three cluster RCTs allocation concealment was not adequate, probably indicative of the non-feasibility of allocation concealment in this type of design due to all participants in one cluster (for example in a practice or with a healthcare provider) being automatically assigned to the same study arm. In three further studies information on random sequence generation or allocation concealment could not be retrieved, leading to a judgment of unclear risk of bias.

We found a high risk of performance bias in 14 of the included studies. In work-directed, psychological, or exercise interventions, blinding of the participant or care provider is not feasible. However, the risk of performance bias also depends on how desirable the intervention is compared to the control group, according to either care providers or participants. One study evaluating a psychological intervention in addition to medication managed to compose two evenly desirable psychological interventions by ensuring an equal number of supportive instead of therapeutic sessions.

In this review, we chose to assess detection and attrition bias separately for sickness absence and depressive symptoms. We felt that not being blind to allocation may bias a self-report assessment of depressive symptoms more than the reporting of a more factual outcome such as the days absent from work in a given period. Also, sickness absence may be retrieved from employee attendance records while depression is measured with a self-report questionnaire. In those instances the lack of blinding of outcome assessment cannot influence the sickness absence but may well bias the depressive outcome.

Potential biases in the review process

This review included studies with a study population of both workers and non-workers. This means that subgroups of the original sample were used for measuring the effect on sickness absence. These studies did not usually present all data for workers separately, but their sickness absence reports were by definition based on the workers in the study population. Some studies included participants with mental disorders other than depression. We included the studies in this review if the authors were willing to provide data for the depressed subgroup.

Subgroup analyses in individual studies may lead to biased results for the following three reasons (Freemantle 2001). First, in the event that no effect is found for the primary outcome it is common that researchers look for a more positive outcome among possible subgroups. Thus the chance for a positive subgroup result would be spuriously increased. Next, in the event of a positive main effect the power for finding an effect in a subgroup would be substantially reduced, and finding no effect for a subgroup could just be a matter of lack of power. Further, testing many subgroups would increase the likelihood of finding a statistically significant result by chance alone. In 19 of the included studies the primary outcome was not work-related and we had to base our conclusions on subgroups of the original sample for which data on work-related outcomes were collected. However, none of the authors specifically looked for a work-related outcome because of the absence of an effect for the primary outcome, and neither did we. Therefore, we are not concerned that this would have influenced our results. This similarly holds for the other argument against the use of subgroups, the testing of multiple groups. We predefined the subgroups; we did not test multiple potential subgroups in the hope of finding a statistically significant group. However, we do agree that a lack of power leading to statistically non-significant findings may have occurred in our review. We were, therefore, careful to not describe

non-significant findings with wide confidence intervals as evidence of no effect.

This review evaluates the effectiveness of a range of interventions aiming to reduce sickness absence in depressed workers rather than one specific intervention. While we believe this is appropriate for a complex and multifactorial outcome such as sickness absence, the categorisation of interventions under the comparisons has been challenging. This categorisation is likely to influence the results as it determines, for each intervention, with which other interventions the results will be pooled and to which other interventions it will be compared. The way interventions are categorised entails a potential bias in the review process.

Another methodological issue concerns the handling of sickness absence data. We accepted both self-report and administrative databases as sources of data on sickness absence. Administrative databases are sometimes considered the gold standard. Agreement between the two sources has been reported to be good (Ferrie 2005; Severens 2000) but also limited (Pole 2006; van Poppel 2002). Furthermore, for the purpose of calculating standardised mean differences (SMDs) we considered sickness absence a construct for which different instruments could be used, as long as they provided information on absenteeism. This meant that as long as we reported SMDs we could incorporate studies with different time spans (and therefore with a different maximum of sickness absence days during follow up) and scales that differed in the maximum score. Also, this enabled us to compare studies from various countries as we know that days of sickness absence tend to be calculated differently in different countries (for instance due to differences in whether calendar days or only work days are included as absenteeism days). Moreover, we transformed reports of days worked into days of sickness absence by extracting the days worked from the days that should have been worked ('the scale maximum'). This is analogous to transforming the scores of a scale in which a high score indicates a good outcome into a scale where a high score indicates a bad outcome. However, for this transformation we had to make inferences about the mean number of hours and the number of hours a day an employee would work in a specific country. In summary, caution is recommended when interpreting sickness absence data in meta-analyses as this is a relatively new field and the methodological issues have not been thoroughly investigated.

Agreements and disagreements with other studies or reviews

A review of the effects of interventions for major depressive disorder on occupational health outcomes was published some years ago (Timbie 2006). Compared to the previous review we were able to identify seven additional studies. Furthermore, Timbie and colleagues chose to combine all studies in their meta-analysis, while we judged the interventions to be too dissimilar for this purpose. Finally, they did not distinguish between time missed from work and workforce participation (both were called labour output). Therefore, we find their conclusion that interventions for depressive disorder have a small but positive effect on labour output too general and believe that we have been able to make better inferences. A more recently published review (Furlan 2012) searched the literature until 2010 and concluded that the evidence was of insufficient quality to determine which interventions are effective and are of value for the management of depression in the workplace. This conclusion was similar to the first published version of this review (Nieuwenhuijsen 2008). This updated version of our review has markedly different conclusions due to the inclusion of a substantially greater number of studies.

AUTHORS' CONCLUSIONS

Implications for practice

We found moderate quality evidence based on three studies that adding a work-directed intervention to a clinical intervention reduces the number of days on sick leave in the medium term when compared to a clinical intervention alone.

There is currently no evidence of a difference in effect on sickness absence of one antidepressant medication compared to another.

We found moderate quality evidence that enhancing primary or occupational care by providing workers with a structured telephone or online cognitive behavioural therapy reduces sickness absence compared to regular care.

We found low quality evidence of no considerable effect for enhanced primary care targeting general practitioners through a quality improvement program.

Based on a single study yielding high quality evidence, we found that a structured telephone outreach and care management program may also lead to reductions in sickness absence of depressed workers.

Implications for research

More research is needed on the addition of work-directed interventions to the clinical care provided. This review shows that such interventions have the potential to reduce sickness absence but the number of studies evaluating these types of interventions is still limited.

More often including occupational outcomes such as sickness absence and work functioning in clinical intervention studies will reveal which clinical interventions can be effective in reducing sickness absence.

To facilitate the synthesis of evidence from various intervention studies, the occupational health field should work towards standardising and validating measures of sickness absence. For future

reviews including absenteeism as an outcome measure, it is advisable to report standardised mean differences instead of means as this takes into account the differences in measurement methods.

ACKNOWLEDGEMENTS

We are grateful to José Luis Fernandez, Martin Keller, Tony Kendrick, Paul Knekt, Paul McCrone, Judith Proudfoot, Renee Romeo, Kathryn Rost, Aart Schene, Gregory Simon, Alan Wade, Penny Bee, Clément Francois, Nicholas Moore, and Ken Wells for kindly providing further information about their studies. We thank the CCDAN group for their support with the first version of this review. We thank Janet Wale for copy editing the text of the updated review. To conclude, we are very grateful for the help of the Cochrane Occupational Safety and Health Review Group and Jani Ruotsalainen in particular.

REFERENCES

References to studies included in this review

Agosti 1991 *{published data only}*
Agosti V, Stewart JW, Quitkin FM. Life satisfaction and psychosocial functioning in chronic depression: effect of acute treatment with antidepressants. *Journal of Affective.Disorders* 1991;**23**(1):35–41.

Bee 2010 *{published data only}*
Bee PE, Bower P, Gilbody S, Lovell K. Improving health and productivity of depressed workers: a pilot randomized controlled trial of telephone cognitive behavioral therapy delivery in workplace settings. *General Hospital Psychiatry* 2010;**32**:337–40.

Burnand 2002 *{published data only}*
* Burnand Y, Andreoli A, Kolatte E, Venturini A, Rosset N. Psychodynamic psychotherapy and clomipramine in the treatment of major depression. *Psychiatric Services* 2002;**53**(5):585–90.

Fantino 2007 *{published data only}*
Fantino B, Moore N, Verdoux H, Auray J. Cost-effectiveness of escitalopram vs. citalopram in major depressive disorder. *Journal of Psychopharmacology* 2007;**22**:107–15.

Fernandez 2005 *{published data only}*
* Fernandez JL, Montgomery S, Francois C. Evaluation of the cost effectiveness of escitalopram versus venlafaxine XR in major depressive disorder. *Pharmacoeconomics* 2005;**23**(2):155–67.

Hees 2013 *{published data only}*
Hees HL, Vries de G, Koeter MW, Schene AH. Adjuvant occupational therapy improves long-term depression recovery and return-to-work in good health in sick-listed employees with major depression: results of a randomised controlled trial. *Occupational and Environmental Medicine* 2013;**70**:252–60.

Hollinghurst 2010 *{published data only}*
Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomized controlled trial. *British Journal of Psychiatry* 2010;**197**:297–304.

Kendrick 2005 *{published data only}*
Kendrick T, Simons L, Mynors-Wallis L, Gray A, Lathlean J, Pickering R, et al. A trial of problem-solving by community mental health nurses for anxiety, depression and life difficulties among general practice patients. The CPN-GP study. *Health Technology Assessment* 2005;**9**(37):1–104.

Knekt 2013 *{published data only}*
Knekt P, Lindfors O, Sares- Jäske L, Virtala E, Härkänen. Randomized trial on the effectiveness of long- and short-term psychotherapy on psychiatric symptoms and working ability during a 5-year follow-up. *Nordic Journal of Psychiatry* 2013;**67**:59–68.

Krogh 2009 *{published data only}*
Krogh J, Saltin B, Gluud C, Nordentoft M. The DEMO trial: a randomized, parallel-group, observer-blinded clinical trial of strength versus aerobic versus relaxation training for patients with mild to moderate depression. *Journal of Clinical Psychiatry* 2009;**70**:790–800.

Krogh 2012 *{published data only}*
Krogh J, Videbech P, Thomsen C, Gluud C, Nordentoft M. DEMO-II trial. Aerobic exercise versus stretching exercise in patients with major depression - a randomized clinical trial. *PLOS ONE* 2012;**7**(10):e48316.

Lerner 2012 *{published data only}*
Lerner D, Adler A, Hermann RC, Hong Chang MS, Ludman EJ, Greenhill A, et al. Impact of a work-focused intervention on the productivity and symptoms of employees with depression. *Journal of Occupational and Environmental Medicine* 2012;**54**:128–35.

McCrone 2004 *{published data only}*
McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: Randomised controlled trial. *British Journal of Psychiatry* 2004;**185**:55–62.

Miller 1998 *{published data only}*
Miller IW, Keitner GI, Schatzberg AF, Klein DN, Thase ME, Rush AJ, et al. The treatment of chronic depression, part 3: psychosocial functioning before and after treatment with sertraline or imipramine. *Journal of Clinical Psychiatry* 1998;**59**(11):608–19.

Noordik 2013 *{published data only}*
Noordik E, Klink JJ van der, Geskus RB, Boer MR de, Dijk FJ van, Nieuwenhuijsen K. Effectiveness of an exposure-

based return-to-work program for workers on sick leave due to common mental disorders: a cluster-randomized controlled trial. *Scandinavian Journal of Work, Environment & Health* 2013;**39**:144–54.

Romeo 2004 ***{published data only}***
Romeo R, Patel A, Knapp M, Thomas C. The cost-effectiveness of mirtazapine versus paroxetine in treating people with depression in primary care. *International Clinical Psychopharmacology* 2004;**19**(3):125–34.

Rost 2004 ***{published data only}***
Rost K, Smith JL, Dickinson M. The effect of improving primary care depression management on employee absenteeism and productivity. A randomized trial. *Medical Care* 2004;**42**(12):1202–10.

Schene 2006 ***{published data only}***
Schene AH, Koeter MWJ, Kikkert MJ, Swinkels JA, Mc Crone P. Adjuvant occupational therapy for work-related major depression works: randomized trial including economic evaluation. *Psychological Medicine* 2006;**2007** (37):3.

Schoenbaum 2001 ***{published data only}***
Schoenbaum M, Unutzer J, Sherbourne C, Duan N, Rubenstein LV, Miranda J, et al. Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. *JAMA* 2001;**286** (11):1325–30.

Simon 1998 ***{published data only}***
Simon GE, Katon W, Rutter C, Von Korff M, Lin E, Robinson P, et al. Impact of improved depression treatment in primary care on daily functioning and disability. *Psychological Medicine* 1998;**28**(3):693–701.

Vlasveld 2013 ***{published data only}***
Vlasveld MC, Feltz-Cornelis CM van der, Adèr HJ, Anema JR, Hoedeman R, Mechelen W van, et al. Collaborative care for sick-listed workers with major depressive disorder: a randomised controlled trial from the Netherlands Depression Initiative aimed at return to work and depressive symptoms. *Occupational and Environmental Medicine* 2013; **70**:223–30.

Wade 2008 ***{published data only}***
Wade AG, Fernandez J, Francois C, Hansen K, Danchenko N, Despiegel N. Escitalopram and duloxetine in major depressive disorder - A pharmacoeconomic comparison using UK cost data. *Pharmacoeconomics* 2008;**26**:969–81.

Wang 2007 ***{published data only}***
Wang PS, Simon SE, Avorn J, Azocar F, Ludman EJ, McCulloch J, et al. Telephone screening, outreach, and care management for depressed workers and impact on clinical and work productivity outcomes. *JAMA* 2007;**298**(12): 1401–11.

References to studies excluded from this review

Aelfers 2013 ***{published data only}***
Aelfers E, Bosma H, Houkes I, Eijk JT van. Effectiveness of a minimal psychological intervention to reduce mild to moderate depression and chronic fatigue in a working population: the design of a randomized trial. *BMC Public Health* 2013;**13**:129.

Ahola 2012 ***{published data only}***
Ahola, K, Vuori J, Toppinen-Tanner S, Mutanen P, Honkonen T. Resrcource-enhancing group intervention against depression at workplace: who benefits? A randomized controlled study with a 7-month follow-up. *Occupational and Environmental Medicine* 2012;**69**:870–6.

Alexopoulos ***{published data only}***
Alexopoulos GS, Raue PJ, Kiosses DN, Mackin RS, Kanellopulos D, McCulloch C, et al. Problem-solving therapy and supportive therapy in older adults with major depression and executive dysfunction. *Archives of General Psychiatry* 2011;**68**(1):33–41.

Amore 2001 ***{published data only}***
Amore M, Jori MC. Faster response on amisulpride 50 mg versus sertraline 50-100 mg in patients with dysthymia or double depression: a randomized, double-blind, parallel group study. *International Clinical Psychopharmacology* 2001;**16**:317–24.

Bakker 2007 ***{published data only}***
Bakker IM, Terluin B, Marwijk HWJ van, Windt DAWM, Rijmen F van der, Mechelen W van, et al. A cluster-randomized trial evaluating an intervention for patients with stress-related mental disorders and sick leave in primary care. *PLOS Clinical Trials* 2007;**26**:1–9.

Barbui 2009 ***{published data only}***
Barbui C, Ostuzzi G, Cipriani A. SSRI plus supportive care more effective than supportive care alone for mild to moderate depression. *Evidence Based Mental Health* 2009; **12**(4):109.

Bech 2000 ***{published data only}***
Bech P, Cialdella P, Haugh MC, Birkett MA, Hours A, Boissel JP, et al. Meta-analysis of randomized controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *British Journal of Psychiatry* 2000;**176**:421–8.

Becker 1998 ***{published data only}***
Becker DR, Drake RE, Bond GR, Xie H, Dain BJ, Harrison K. Job terminations among persons with severe mental illness. *Community Mental Health Journal* 1998;**34**:71–82.

Blonk 2007 ***{published data only}***
Blonk RWB, Brenninkmeijer V, Lagerveld SE, Houtman ILD. Return to work: a comparison of two cognitive behavioural interventions in cases of work-related psychological complaints among the self-employed. *Work & Stress* 2006;**20**:129–44.

Boyer 1998 ***{published data only}***
Boyer P, Danion JM, Bisserbe JC, Hotton JM, Troy S. Clinical and economic comparison of sertraline and fluoxetine in the treatment of depression. A 6-month double-blind study in a primary-care setting in France. *Pharmacoeconomics* 1998;**13**(1 Pt 2):157–69.

Brandes 2011 *{published data only}*
Brandes VMT. Efficacy of auditory stimulation programs for the treatment of depression, dysthymia and symptoms of burnout - RCT results. European Psychiatry conference. 2011.

Brouwers 2007 *{published data only}*
Brouwers EPM, Bruijne MC de, Terluin B, Tiemens BG, Verhaak PFM. Cost-effectiveness of an activating intervention by social workers for patients with minor mental disorders on sick leave: a randomized controlled trial. *The European Journal of Public Health* 2007;**17**(2): 214–20.

Carlin 2010 *{published data only}*
Carlin E. The effect of a motivational interviewing style in cognitive therapy for depression [Dissertation]. http://gradworks.umi.com/34/53/3453319.html 2010.

Castillo-Pérez 2010 *{published data only}*
Castillo-Pérez S, Gómez-Pérez V, Calvillo Velasco M, Pérez-Campos E, Mayoral M. Effects of music therapy on depression compared with psychotherapy. *The Arts in Psychotherapy* 2010;**37**:387–90.

Dick 1985 *{published data only}*
Dick P, Cameron L, Cohen D, Barlow M. Day and full time psychiatric treatment. *British Journal of Psychiatry* 1985; **147**:246–9.

Dunlop 2011 *{published data only}*
Dunlop BW, Reddy S, Yang L, Lubaczewski S, Focht K, Guico-Pabia CJ. Symptomatic and functional improvement in employed depressed patients. *Journal of Clinical Psychopharmacology* 2011;**31**:569–76.

Eklund 2012 *{published data only}*
Eklund M, Wästberg BA, Erlandsson L. Work outcomes and their predictors in the Redesigning Daily Occupations (ReDO) rehabilitation programme for women with stress-related disorders. *Australian Occupational Therapy Journal* 2013;**60**:85–92.

Erkkilä 2011 *{published data only}*
Erkkilä J, Punkanen M, Fachner J, Ala-Ruona E, Pöntiö I, Tervaniemi M, et al. Individual music therapy for depression: randomized controlled trial. *British Journal of Psychiatry* 2011;**199**:132–39.

Finley 2003 *{published data only}*
Finley PR, Rens HR, Pont JT, Gess SL, Louie CL, Bull SA, et al. Impact of a collaborative care model on depression in a primary care setting: a randomized controlled trial. *Pharmocotherapy* 2003;**23**:1175–85.

Folke 2012 *{published data only}*
Folke F, Parling T, Melin L. Acceptance and commitment therapy for depression: a preliminary randomized clinical trial for unemployed on long-term sick leave. *Cognitive and Behavioural Practice* 2012;**19**:583–94.

Forman 2012 *{published data only}*
Forman EM, Shaw JA, Goetter EM, Herbert JD, Park JA, Yuen EK. Long-term follow up of a randomized controlled trial comparing acceptance and commitment therapy and standard cognitive behaviour therapy for anxiety and depression. *Behavior Therapy* 2012;**43**:801–11.

Furukawa 2012 *{published data only}*
Furukawa TA, Horikoshi M, Kawakami N, Kadota M, Sasaki M, Sekiya Y, et al. Telephone cognitive-behavioural therapy for subthreshold depression and presenteeism in workplace: a randomized controlled trial. *PLOS ONE* 2012;**7**:1–9.

Gournay 1995 *{published data only}*
Gournay K, Brooking J. The community psychiatric nurse in primary care: an economic analysis. *Journal of Advanced Nursing* 1995;**22**(4):769–78.

Hackett 1987 *{published data only}*
Hackett GI, Boddie HG, Harrison P. Chronic muscle contraction: the importance of depression and anxiety. *Journal of the Royal Society of Medicine* 1987;**80**:689–91.

Hirani 2010 *{published data only}*
Hirani SS, Karmaliani R, McFarlane J, Asad N, Madhani F, Shehzad S. Testing a community derived intervention to promote women's health: preliminary results of a 3-arm randomized controlled trial in Karachi, Pakistan. *Southern Online Journal of Nursing Research* 2010;**10**(3):n.a.

Hordern 1964 *{published data only}*
Hordern A, Burt CG, Gordon WF, Holt NF. Amitriptyline in depressive states: Six-month treatment results. *British Journal of Psychiatry* 1964;**110**:641–7.

Knekt 2011 *{published data only}*
Knekt P, Lindfors O, Laaksonen MA, Renlund C, Haaramo P, Härkänen, et al. Quasi-experimental study on the effectiveness of psychoanalysis, long-term and short-term psychotherapy on psychiatric symptoms, work ability and functional capacity during a 5-year follow-up. *Journal of Affective Disorders* 2011;**132**:37–47.

Kojima 2010 *{published data only}*
Kojima R, Fujisawa D, Tajima M, Shibaoka M, Kakinuma M, Shima S, et al. Efficacy of cognitive behavioral therapy training using brief e-mail sessions in the workplace: a controlled clinical trial. *Industrial Health* 2010;**48**:495–502.

Kroenke 2001 *{published data only}*
Kroenke K, West SL, Swindle R, Gilsenan A, Eckert GJ, Dolor R, et al. Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care - a randomized trial. *JAMA* 2001;**286**:2947–55.

Kuhs 1996 *{published data only}*
Kuhs H, Färber D, Borgstädt S, Mrosek S, Tölle R. Amitriptyline in combination with repeated late deep deprivation versus amitriptyline alone in major depression. A randomized study. *Journal of Affective Disorders* 1996;**37**: 31–41.

Lagerveld 2012 *{published data only}*
Lagerveld SE, Blonk RWB, Brenninkmeijer V, Wijngaards-de Meij L, Schaufeli WB. Work-focused treatment of common mental disorders and return to work: a comparative outcome study. *Journal of Occupational Psychology* 2012;**17**:220–34.

Lam 2012 *{published data only}*
Lam RW, McLeod J, Woo C, Michalak EE, Yatham LN, Bond DJ. A systematic review of the treatment effects of antidepressants on occupational functioning. Presented at the International Society for Affective Disorders. UK, 2012.

Lexis 2011 *{published data only}*
Lexis MAS, Jansen NWH, Huibers MJH, Amelsvoort LGPM van, Berkouwer A, Tjin A, Ton G, et al. Prevention of long-term sickness absence and major depression in high-risk employees: a randomised controlled trial. *Occupational and Environmental Medicine* 2011;**68**:400–7.

Martinez 2011 *{published data only}*
Martinez JM, Katon W, Greist JH, Kroenke K, Thase ME, Meyers AL, et al. A pragmatic 12-week, randomized trial of duloxetine versus generic selective serotonin-reuptake inhibitors in the treatment of adult outpatients in a moderate-to-severe depressive episode. *International Clinical Psychopharmacology* 2012;**27**:17–26.

Meyer 2009 *{published data only}*
Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. *Journal of Medical Internet Research* 2009;**11**:e15.

Mino 2006 *{published data only}*
Mino Y, Babazono A, Tsuda T, Yasuda N. Can stress management at the workplace prevent depression? A randomized controlled trial. *Psychotherapy and Psychosomatics* 2006;**75**:177–82.

Morgan 2011 *{published data only}*
Morgan AJ, Form AF, Mackinnon AJ. Protocol for a randomized controlled trial investigating self-help email messages for sub-treshold depression: the Mood Memos study. *Trials* 2011;**12**:1–9.

Mundt 2001 *{published data only}*
Mundt JC, Clarke GN, Burroughs D, Brenneman DO, Griest JH. Effectiveness of antidepressant pharmacotherapy: the impact of medication compliance and patient education. *Depression and Anxiety* 2001;**13**:1–10.

Oakes 2012 *{published data only}*
Oakes TMM, Meyers AL, Marangell LB, Ahl J, Prakash A, Thase ME, Kornstein SG. Assessment of depressive symptoms and functional outcomes in patients with major depressive disorder treated with duloxetine versus placebo. *Human Psychopharmacology: Clinical and Experimental* 2012;**27**:47–56.

Salminen 2008 *{published data only}*
Salminen JK, Karlsson H, Hietala J, Kajander J, Aalto S, Markkula J, et al. Short-term psychodynamic psychotherapy and fluoxetine in major depressive disorder: a randomized comparative study. *Psychotherapy and Psychosomatics* 2008; **77**:351–7.

Sandahl 2011 *{published data only}*
Sandahl C, Lundberg U, Lindgren A, Rylander G, Herlofsen J, Nygren ÅN. Two forms of group therapy and individual treatment of work-related depression: a one-year follow-up study. *International Journal of Group Psychotherapy* 2011;**61**: 539–55.

Schmitt 2008 *{published data only}*
Schmitt AB, Volz HP, Wiedemann K, Fritze J, Loeschmann PA. Therapy effects on working ability by venlafaxine compared to serotonin reuptake inhibitors in the treatment of major depressive disorder of different severity. *Gesundheitsökonomie & Qualitätsmanagement* 2008;**13**: 269–75.

Schoenbaum 2002 *{published data only}*
Schoenbaum M, Unutzer J, McCaffrey D, Duan N, Sherbourne C, Wells KB. The effects of primary care depression treatment on patients' clinical status and employment. *Health Services Research* 2002;**37**(5):1145–58.

Simon 2000 *{published data only}*
Simon GE, Revicki D, Heiligenstein J, Grothaus L, Von Korff M, Katon WJ, et al. Recovery from depression, work productivity, and health care costs among primary care patients. *General Hospital Psychiatry* 2000;**22**(3):153–62.

Sir 2005 *{published data only}*
Sir A, D'Souza RF, Uguz S, George T, Psych MRC, Vahip S, et al. Randomized trial of sertraline versus venlafaxine XR in major depression: efficacy and discontinuation symptoms. *Journal of Clinical Psychiatry* 2005;**66**:1312–20.

Stant 2009 *{published data only}*
Stant AD, TenVergert EM, Kluiter H, Conradi HJ, Smit A, Ormel J. Cost-effectiveness of a psychoeducational relapse prevention program for depression in primary care. *The Journal of Mental Health Policy and Economics* 2009;**12**: 195–204.

Wells 2000 *{published data only}*
Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unutzer J, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA* 2000; **283**(2):212–20.

Zambori 2002 *{published data only}*
Zambori J, Szadoczky E, Rozsa S, Furedi J. Cost-outcome of anxiety treatment intervention in primary care in Hungary. *The Journal of Mental Health Policy and Economics* 2002;**5** (3):115–20.

Zeeuw 2010 *{published data only}*
Zeeuw ELEJ de, Tak ECPM, Dusseldorp E, Hendriksen IJM. Workplace exercise intervention to prevent depression: a pilot randomized controlled trial. *Mental Health and Physical Activity* 2010;**3**:72–7.

References to ongoing studies

Beurden 2013 *{published data only}*
Beurden KM, Brouwers EP, Joosen MC, Terluin B, Klink JJ, Weeghel J. Effectiveness of guideline-based care by occupational physicians on the return-to-work of workers with common mental disorders: design of a cluster-randomized controlled trial. *BMC Public Health* 2013;**13**: 1–8.

Geraedts 2013 *{published data only}*
Geraedts AS, Kleiboer AM, Wiezer NM, Mechelen W van, Cuijpers P. Web-based guided self-help for employees with depressive symptoms (Happy@Work): design of a randomized controlled trial. *BMC Psychiatry* 2013;**13**:1–10.

Heer 2013 *{published data only}*
Heer EW de, Dekker J, Van Eck-Sluijs JF van der, Beekman ATF, Marwijk HWJ van, Holwerda TJ, et al. Effectiveness and cost-effectiveness of transmural collaborative care with consultation letter (TCCCL) and duloxetine for major depressive disorder (MDD) and (sub)chronic pain in collaboration with primary care: design of a randomized placebo-controlled muli-Centre trial: TCC: PAINDIP. *BMC Psychiatry* 2013;**13**:1–14.

Hellstrom 2013 *{published data only}*
Hellstrom L, Bech P, Nordentoft M, Lindschou J, Eplov LF. The effect of IPS-modified, an early intervention for people with mood and anxiety disorders: study protocol for a randomized clinical superiority trial. *Trials* 2013;**14**:1–10.

Warmerdam 2007 *{published data only}*
Warmerdam L, Straten A van, Cuijpers P. Internet-based treatment for adults with depressive symptoms. *BMC Psychiatry* 2007;**7**:1–7.

Additional references

Adler 2006
Adler DA, McLaughlin TJ, Rogers WH, Chang H, Lapitsky L, Lerner D. Job performance deficits due to depression. *American.Journal of.Psychiatry* 2006;**163**(9):1569–76.

APA 1994
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. 4th Edition. Washington, DC: American Psychiatric Association, 1994.

APA 2010
American Psychiatric Association (APA). Practice guideline for the treatment of patients with major depressive disorder. 3rd edition. American Journal of Psychiatry. Arlington (VA): American Psychiatric Association (APA), 2010:152.

APA 2013
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Arlington, VA: American Psychiatric Association, 2013.

Beck 1987
Beck AT, Steer RA. *BDI: Beck Depression Inventory Manual*. New York, NY: Psychological Corporation, 1987.

Bowling 1995
Bowling A. What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Social Science and Medicine* 1995;**41** (10):1447–62.

Brown 2013
Brown MM, Brown GC. Update on value-based medicine. *Current Opinion in Ophthalmology* 2013;**24**(3):183–9.

Burton 2004
Burton WN, Pransky G, Conti DJ, Chen CY, Edington DW. The association of medical conditions and presenteeism. *Journal of Occupational and Environmental Medicine* 2004; **46**(6 Suppl):S38–S45.

Cooney 2013
Cooney GM, Dwan K, Greig CA, Lawlor DA, Rimer J, Waugh FR, et al. Exercise for depression. *Cochrane Database of Systematic Reviews* 2013;**Sep 12;9:CD004366**: doi: 10.1002/14651858.CD004366.pub6.

Dewa 2000
Dewa CS, Lin E. Chronic physical illness, psychiatric disorder and disability in the workplace. *Social Science and Medicine* 2000;**51**(1):41–50.

Donner 2002
Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Statistics in Medicine* 2002;**21**:2971–80.

Endicott 1997
Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. *Psychopharmacology Bulletin* 1997;**33**(1):13–6.

Ferrari 2013
Ferrari AJ, Charlson FJ, Norman RE, Flaxman AD, Patten SB, Vos T, et al. The epidemiological modelling of major depressive disorder: application for the Global Burden of Disease Study 2010. *PLoS One* 2013;**8**(7):e69637. doi: 10.1371/journal.pone.0069637.

Ferrie 2005
Ferrie JE, Kivimaki M, Head J, Shipley MJ, Vahtera J, Marmot MG. A comparison of self-reported sickness absence with absences recorded in employers' registers: evidence from the Whitehall II study. *Occupational and Environmental Medicine* 2005;**62**(2):74–9.

Freemantle 2001
Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic?. *BMJ* 2001;**322**(7292): 989–91.

Furlan 2012
Furlan AD, Gnam WH, Carnide N, Irvin E, Amick BC 3rd, DeRango K, et al. Systematic review of intervention practices for depression in the workplace. *Journal of Occupational Rehabilitation* 2012;**22**(3):312–21.

GRADEpro 2008 [Computer program]
Brozek J, Oxman A, Schünemann H. GRADEpro. Version 3.2 for Windows. GRADE working group, 2008.

Haafkens 2006
Haafkens J, Moerman C, Schuring M, van Dijk F. Searching bibliographic databases for literature on chronic disease and work participation. *Occupational Medicine* 2006;**56**(1): 39–45.

Hamilton 1967
Hamilton M. Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology* 1967;**6**(4):278–96.

Haslam 2005
Haslam C, Atkinson S, Brown S, Haslam RA. Perceptions of the impact of depression and anxiety and the medication for these conditions on safety in the workplace. *Occupational and Environmental Medicine* 2005;**62**(8):538–45.

Hees 2013b
Hees HL, Koeter MW, Schene AH. Longitudinal relationship between depressive symptoms and work outcomes in clinically treated patients with long-term sickness absence related to major depressive disorder. *Journal of Affective Disorders* 2013;**148**(2-3):272–7.

Higgins 2011
Higgins JPT, Green S (editors). The Cochrane Collaboration, 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, 2011*. Available from www.cochrane-handbook.org, 2011.

Hirschfeld 2000
Hirschfeld RM, Montgomery SA, Keller MB, Kasper S, Schatzberg AF, Moller HJ, et al. Social functioning in depression: a review. *Journal of Clinical Psychiatry* 2000;**61**(4):268–75.

Kessler 2006
Kessler RC, Akiskal HS, Ames M, Birnbaum H, Greenberg P, Hirschfeld RM, et al. Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers. *American Journal of Psychiatry* 2006;**163**(9):1561–8.

Kessler 2009
Kessler RC, Aguilar-Gaxioloa S, Alonso J, Chatterji S, Lee S, Ormel J, et al. The global burden of mental disorders: An update from the WHO World Mental Health (WMH) Suveys. *Epidemiologia e Psichiatria Sociale* 2009;**18**(1):23–33.

Lagerveld 2010
Lagerveld SE, Bültmann U, Franche RL, van Dijk FJ, Vlasveld MC, van der Feltz-Cornelis CM, et al. Factors associated with work participation and work functioning in depressed workers: a systematic review. *Journal of Occupational Rehabilitation* 2010;**20**(3):275–92.

Lerner 2008
Lerner D, Henke RM. What does research tell us about depression, job performance, and work productivity?. *Journal of Occupational and Environmental Medicine* 2008;**50**(4):401–10.

Montgomery 1979
Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* 1979;**134**:382-9.

Murray 2012
Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;**380**(9859):2197–223.

NICE 2010
National Collaborating Centre for Mental Health. The treatment and management of depression in adults (updated edition). *http://www.nice.org.uk/nicemedia/live/12329/45896/45896.pdf.* Leicester & London: The British Psychological Society and The Royal College of Psychiatrists, 2010.

Nieuwenhuijsen 2010
Nieuwenhuijsen K, Franche RL, van Dijk FJH. Work functioning measurement: tools for occupational mental health research. *Journal of Occupational and Environmental Medicine* 2010;**52**(8):778–90.

Pole 2006
Pole JD, Franche RL, Hogg-Johnson S, Vidmar M, Krause N. Duration of work disability: a comparison of self-report and administrative data. *American Journal of Industrial Medicine* 2006;**49**(5):394–401.

RevMan 2012 [Computer program]
The Nordic Cochrane Centre. The Cochrane Collaboration. Review Manager (RevMan). Version 5.2. Copenhagen: The Nordic Cochrane Centre. The Cochrane Collaboration, 2012.

Severens 2000
Severens JL, Mulder J, Laheij RJ, Verbeek AL. Precision and accuracy in measuring absence from work as a basis for calculating productivity costs in The Netherlands. *Social Science and Medicine* 2000;**51**(2):243–9.

Sheehan 1996
Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. *International Clinical Psychopharmacology* 1996;**11 Suppl 3**:89–95.

Spitzer 1979
Spitzer RL, Endicott J, Williams JB. Research diagnostic criteria. *Archive of General Psychiatry* 1979;**36**(12):1381–3.

Suzuki 2004
Suzuki K, Ohida T, Kaneita Y, Yokoyama E, Miyake T, Harano S, et al. Mental health status, shift work, and occupational accidents among hospital nurses in Japan. *Journal of Occupational Health* 2004;**46**(6):448-54.

Timbie 2006
Timbie JW, Horvitz-Lennon M, Frank RG, Normand SL. A meta-analysis of labor supply effects of interventions for major depressive disorder. *Psychiatric Services* 2006;**57**(2):212–8.

van Poppel 2002
van Poppel MN, de Vet HC, Koes BW, Smid T, Bouter LM. Measuring sick leave: a comparison of self-reported data on sick leave and data from company records. *Occupational Medicine* 2002;**52**(8):485–90.

Verbeek 2005
Verbeek J, Salmi J, Pasternack I, Jauhiainen M, Laamanen I, Schaafsma F, et al. A search strategy for occupational health intervention studies. *Occupational and Environmental Medicine* 2005;**62**(10):682–7.

Wells 1989

Wells KB, Stewart A, Hays RD, Burnam MA, Rogers W, Daniels M, et al. The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *JAMA* 1989;**262**(7):914–9.

WHO 1992

World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders*. Geneva: World Health Organization, 1992.

References to other published versions of this review

Nieuwenhuijsen 2008

Nieuwenhuijsen K, Bültmann U, Neumeyer-Gromen A, Verhoeven AC, Verbeek JH, Feltz-Cornelis CM. Interventions to improve occupational health in depressed people. *Cochrane Database of Systematic Reviews* 2008, Issue 2. [DOI: 10.1002/14651858.CD006237.pub2

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Agosti 1991

Methods	Double-blind randomised trial with four arms (3 treatment and one placebo). Recruitment: unclear. Follow up: 6 weeks. Lost to follow up: 29.5%
Participants	61 were randomised (T1: 38, C: 23). Setting: Outpatients in New York, USA Inclusion: - DSM-III diagnosis of depressive disorder - mood reactivity (i.e. significant lifting of mood in response to positive environmental events) - onset prior to age 21 yrs - rated by experienced clinician to be depressed for most or virtually all of the time through adulthood Mean age: 35 yrs (SD 8.9) Female: 52% Single: 57% Married: 23% Divorced or separated: 19.6% Working: 70%
Interventions	T1: Treatment with increasing dose of either TCA or MAO - 60 to 90 mg/day of phenelzine (T1a) - 200 to 3000 mg/day of imipramine (T1b) - 40 mg/day of L-deprenyl (T1c) Duration: 6 weeks. C: 4 to 6 placebo pills/day. Duration: 6 weeks
Outcomes	Absenteeism: 1) hours worked in past week (baseline and at 6 weeks) Clinical: 1) CGI (measured but not reported!) 2) HAM-D (measured but not reported!) Productivity: 1) work functioning of the LIFE scale (psychosocial functioning part) (baseline and at 6 weeks)
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Random sequence generation not reported "Following baseline evaluation, patients were treated with single-blind placebo for 1-2 weeks, those who were still depressed

		were randomly assigned to 6 weeks of treatment with increasing doses of one of four agents in a double blind design."
Allocation concealment (selection bias)	Unclear risk	Allocation concealment not reported "Following baseline evaluation, patients were treated with single-blind placebo for 1-2 weeks, those who were still depressed were randomly assigned to 6 weeks of treatment with increasing doses of one of four agents in a double blind design."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	A double blind design was used "Following baseline evaluation, patients were treated with single-blind placebo for 1-2 weeks, those who were still depressed were randomly assigned to 6 weeks of treatment with increasing doses of one of four agents in a double blind design."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Patients reported sick leave in an interview, but were blinded to treatment allocation "Sick leave was assessed by the LIFE. The LIFE is a semi-structured interview which tracks episodes of psychiatric illness. The portion of the LIFE which we used assessed the psychosocial functioning during the week in five areas; employment..etc. The LIFE was administered to the patient by the treating physician."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	Depressive symptoms were determined by personnel, were blinded to treatment allocation "Clinical outcome was determined by the treating psychiatrist on the basis of Clinical Global Improvement."
Incomplete outcome data (attrition bias) Depressive symptoms	Unclear risk	Outcome not reported
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up is considered to be high: T1: 28.9%; T2: 30.4%, even though the proportion of incomplete data was comparable in both groups
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk

Agosti 1991 (*Continued*)

Other bias	Unclear risk	None Identified

Bee 2010

Methods	RCT. Recruitment: over 10 months, human resources mailed all potential participants a study information pack. Follow up: 3 months. Lost to follow up: overall 40%, subgroup depressed workers: 0%
Participants	53 were randomised (T1: 26; T2: 27). Subgroup of depressed workers: 12. Setting: large communications company. Inclusion: employees of a large communications company absent from work with mild to moderate mental health difficulties for 8 to 90 days authorised by general practitioner certificate Exclusion: severe or complex disorders (psychosis, comorbid personality disorder), degenerative cognitive disorders, substance misuse or active self-harm For the subgroup of depressed workers: mean age: 50.9 (SD 10.04) male: 58%
Interventions	T1: Telephone CBT, delivered over 12 weeks by one of two registered graduate mental health workers. Participants worked with therapists through regular phone calls to identify and challenge negative thoughts, develop self-care skills and complete workbook exercises emphasizing behavioural activation. Therapists received 12 h of didactic instruction and role play and weekly supervision from a senior CBT therapist T2: Usual care, primary and occupational health services.
Outcomes	Absenteeism: 1) self-reported actual working hours (HPQ) in last four weeks Clinical: 1) depression, assessed by the HADS
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Personal communication: "Yes there was a random component in the sequence generation - and the sequence was held by an independent trial units."
Allocation concealment (selection bias)	Low risk	"Randomization was conducted centrally by an independent service, with minimization on age, gender and illness severity". "[. ..] internal validity was heightened trough allocation concealment via central randomization [..]"

Bee 2010 (*Continued*)

Blinding of participants and personnel (performance bias) Sick Leave	High risk	Due to the nature of the intervention, the participants could not be blinded
Blinding of outcome assessment (detection bias) Sick Leave	High risk	The actual working hours were assessed by the participants themselves. As they were aware of the allocation status, risk of detection bias is considered to be high
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression is assessed by the HADS, which is a self-reported instrument. As the participants were aware of their allocation status, risk of detection bias is considered to be high
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Personal communication: "For the subgroup of depressed workers, there is no loss to follow up."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Personal communication: "For the subgroup of depressed workers, there is no loss to follow up."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Burnand 2002

Methods	RCT, random assignment stratified by presence of personality disorder, past major depressive syndrome and gender; two conditions. Recruitment: screening by nurse and psychiatrist of consecutive patients referred for acute outpatient treatment. Follow up: 10 weeks. Lost to follow up: 22%
Participants	95 were randomised (T1: 35; C: 39); Setting: outpatient community mental health centre in Switzerland; Inclusion: age 20 to 65 years, new episode of care, MDD DSM-IV (SCID) + HDRS at least 20; Exclusion: bipolar disorder, psychotic symptoms, severe substance dependence, organic disorder, mental retardation, history of severe intolerance to clomipramine, poor command of French language Age: T1: 36 (SD 9.5); C: 36.7 (SD 10.4) Female: T1: 66%; C: 56% Stable employment: T1: 71%; C: 82%
Interventions	T1: Psychodynamic psychotherapy: individual sessions by nurse + clomipramine: 25 mg first day, gradually increasing to 125 mg on fifth day (dosage adjustment allowed). Refusal or severe side effects: 20 to 40 mg citalopram per day. Duration: 10-week program, frequency psychotherapy sessions not fixed, duration of clomipramine 10 weeks

Burnand 2002 (*Continued*)

	C: Supportive care: individual sessions: empathic listening, guidance and support. + clomipramine: 25 mg first day, gradually increasing to 125 mg by fifth day (dosage adjustment allowed). Refusal or severe side effects: 20 to 40 mg citalopram per day. Duration supportive care: not fixed, duration clomipramine 10 weeks
Outcomes	Absenteeism: 1) number of days of sick leave in 10 weeks Clinical: 1) full remission (at most 7 HDRS) (at 10 weeks) 2) severity of depression (HDRS score; GAS) (at 10 weeks)
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Randomisation procedure not reported
Allocation concealment (selection bias)	Unclear risk	Randomisation procedure not reported
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	No blinding but risk of performance bias low as both treatments can be considered equally desirable for patients "Both treatments involved the same clomipramine protocol and intensive nursing in a specialized milieu. In addition, the amount of structured psychodynamic psychotherapy provided during combined treatment was comparable to the amount of supportive care provided during treatment with clomipramine alone."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Outcome assessor for sick leave was blinded, but (non-blinded) patients had to report the number of sick leave days to them "The psychologists who made the assessments of hospitalizations, number of days of sick leave, and GAS scores were blinded to each patient's treatment assignment."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	"The individuals who rated the presence and severity of major depression and HSRS scores at ten weeks were not blinded to treatment assignment."

Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up is high: 22%. Risk of attrition bias due to follow up losses is therefore considered to be high, although multiple analyses were used to study the effect on the findings and the authors conclude otherwise: "Twenty-one patients (12 in the experimental and nine in the control group, or 22 percent) were excluded from the analysis--four who did not return for treatment (three in the experimental group and one in the control group), three who dropped out against medical advice (two in the experimental group and one in the control group) , and 14 who were discharged because they had exclusion characteristics that were not detected at entry, including severe alcohol or drug dependence (five in each group) and adverse effects (two in each group). These patients were not significantly different from the other patients in terms of the main outcome variables at intake. The 74 patients who completed the study were not significantly different from the 21 who were withdrawn or from the group of 95 as a whole. To control for intent to treat, the analyses were repeated with all 95 patients who had been randomly assigned to treatment." "This finding was unchanged when we repeated the analyses and controlled for age, gender, initial severity of depression, GAS score at intake, compliance and intent to treat"
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up is high: 22%. Risk of attrition bias due to follow up losses is therefore considered to be high, although multiple analyses were used to study the effect on the findings and the authors conclude otherwise: "Twenty-one patients (12 in the experimental and nine in the control group, or 22 percent) were excluded from the analysis--four who did not return for treatment (three in the experimental group and one in the control group), three who dropped out against medical advice (two in the experimental group and one in the control group) , and 14 who were discharged because they had exclusion characteristics that were not

Burnand 2002 (*Continued*)

		detected at entry, including severe alcohol or drug dependence (five in each group) and adverse effects (two in each group). These patients were not significantly different from the other patients in terms of the main outcome variables at intake. The 74 patients who completed the study were not significantly different from the 21 who were withdrawn or from the group of 95 as a whole. To control for intent to treat, the analyses were repeated with all 95 patients who had been randomly assigned to treatment." "This finding was unchanged when we repeated the analyses and controlled for age, gender, initial severity of depression, GAS score at intake, compliance and intent to treat"
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None Identified

Fantino 2007

Methods	RCT. Recruitment: patients were recruited by psychiatrists or by general practitioners. Follow up: 8 weeks. Lost to follow up: 8.1%
Participants	280 were randomised (T1: 138; T2: 142). Setting: outpatient; general or psychiatric practices in France. Inclusion: all patients fulfilling the DSM-IV criteria for MDD and having a baseline MADRS total score of at least 30 were eligible for the study. Exclusion: patients meeting DSM-IV for primary diagnoses for any axis I disorder other than MDD or those with a history of mania, bipolar disorder, schizophrenia or other psychotic disorder, obsessive-compulsive disorder, cognitive disorder including mental retardation or personality disorder, patients who met the DMS-IV criteria for substance abuse or dependence within the past 12 months, or used a depot antipsychotic within 6 months before study inclusion or any antipsychotic or anticonvulsant medications within 2 weeks before the first administration of study medication Male: T1: 28.3%; T2: 38.0% Age: T1: 44.1 (SD 10.9); T2: 46.2 (SD 11.1) Family situation: T1: 23.9% single; T2: 16.2% single T1: 49.3% married, living with partner; T2: 50.7% married living with partner T1: 26.8% separated, divorced, widowed; T2: 33.1% separated, divorced, widowed Occupational status: T1: 35.5% unemployed; T2: 29.6% unemployed T1: 64.5% employed; T2: 70.4% T1: 4.5% craftsman, tradesman; T2: 7.0% craftsman, tradesman

	T1: 9.0% manager; T2: 12.0% manager T1: 21.3% technician; T2: 30.0% technician T1: 9.0% workman; T2: 4.0% workman
Interventions	T1: Escitalopram (SSRI) 10 mg daily during the first week, 20 mg per day for the remaining 7 weeks T2: Citalopram (SSRI) 20 mg/day daily during the first week, 40 mg per day for the remaining 7 weeks All study medications were provided in identical blister packs of identical capsules administered as one capsule per day, regardless of dose or treatment group. No adjustment of dosage was allowed
Outcomes	Absenteeism: 1) says of sick leave for the 2-month pre-study period and for the 8-week study period (percentage of patients and mean consumption of those patients) Clinical: 1) sepression severity, assessed by the Montgomery-Asberg Depression Scale (MADRS) 2) remission, defined as the total score MADRS of $\leq$ 12 3) MADRS-S, the self-reported version of MADRS
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Personal communication: "Allocation was random. This includes random allocation using equal block sizes."
Allocation concealment (selection bias)	Low risk	Personal communication: "Allocation was concealed. Investigators allotted patients to a treatment defined by the patient inclusion number. All treatments were prepared and identical, the only difference being the treatment number, corresponding to the allocation table, which was kept by the person who prepared the treatments. The investigators were not aware of the nature of the treatments."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Trial is double-blind: "Those meeting the eligibility criteria were randomly assigned to receive double-blind, fixed doses of either escitalopram 20 mg daily or citalopram 40 mg daily during 8 weeks, with equal block randomization at baseline." "All study medications were provided in identical blister packs of identical capsules

		administered as one capsule per day, regardless of dose or treatment group." Personal communication: "The psychiatrist or GP both included the patient, dispensed the study medication, and did the assessments. Patient and investigator were both blind to the treatment, which were identical in aspect. Since this was not placebo-controlled, both comparators were active and quite similar, differing only be the presence of 20 mg R-citalopram in the 40 mg citalopram. This actually reduces the risk of unblinding by recognizable drug effects or side-effects."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	"A standardized form was used by trained investigators to record healthcare services and days of sick leave for the 2-month pre-study period and for the 8-week study period." Since the investigators were blinded, the risk of bias is considered to be low
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	The MADSR was done by investigators who are trained or confirmed in the proper use of the MADSR scores and who were blinded for the allocation status. The MADSR-S is a self-reported version, but patients were also blinded for treatment allocation
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Loss to follow up is considered to be low. T1: 4.3%; T2: 10.6%
Incomplete outcome data (attrition bias) Sick Leave	Low risk	No missing sick leave data: "Valid resource utilization information corresponding to the pre study and study periods was thus available for 280 patients."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Fernandez 2005

Methods	Randomised, double-blind, flexible-dose, multinational, clinical trial with a one-week run-in period with no treatment. After randomisation: two treatment arms Recruitment: patient were asked to participate by GP. Follow up: 8 weeks. Lost to follow up: 16%
Participants	293 were randomised (T1: 148; T2: 145). Setting: primary care at 44 sites in 8 European countries. Inclusion: patients in primary care Age 18 to 85 yrs, DSM-IV diagnosis of MDD (current or first), Minimal MADRS score of 18. Exclusion: History of mania or any bipolar disorder, schizophrenia or any psychotic disorder, Currently suffering from obsessive-compulsive disorder, eating disorder, mental retardation, any pervasive development disorder, or cognitive disorder (DSM-IV criteria), MADRS of at least 5 on item 10 (suicidal thoughts), Alcohol or drug abuse problems within the previous 12 months, Having had treatment with: antipsychotics, antidepressants, psychotropics (except zolpidem or stable low doses of benzodiazepines for insomnia), serotonin receptor antagonists, lithium, carbamazepine, valproate, or valpromide, ECT, treatment with CBT or psychotherapy, Being pregnant or breastfeeding, Medications likely to interfere with the study Mean age T1: 48.4; T2: 46.5 Sex: T1: 75.4% female; T2: 71.2% female Married or cohabiting: T1: 61.9%; T2: 56% Employed: T1: 51.5%; T2: 60% Long-term sickness absence: T1: 11.1%; T2: 11.2% Higher education: T1: 9.5%. T2: 11.2%
Interventions	T1: Escitalopram (SSRI): initial 10 mg/day. At week 2 or 4 dose could be increased to 20 mg/day at the investigator's discretion if patient's response was unsatisfactory. After 8 weeks of treatment, 1 week run-out period. Patients on 20 mg/day were down-tapered to 10 mg for the first 4 days and placebo the last 3. Patients on lower dose received 7 days of placebo T2: Venlafaxine XR (SNRI), initially 75 mg/day. At week 2 or 4 dose could be increased to 150 mg/day at the investigator's discretion if patient's response was unsatisfactorily. After 8 weeks of treatment, 1-week run-out period. Patients on 150 mg/day were down-tapered to 75 mg for the first 4 days and placebo the last 3. Patients on lower dose received 7 days of placebo
Outcomes	Absenteeism: 1) % of patients on sick leave and average length of sick leave per week (3 months prior baseline and during 8 weeks of study) 2) personal communication; days of sick leave during 8 weeks of study, for workers only Clinical: 1) MADRS (at 8 weeks) 2) HAM-D (at 8 weeks)
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**

Random sequence generation (selection bias)	Low risk	Personal communication with first author: "Patients who met the selection criteria at the baseline visit were assigned to 8 weeks of double-blind treatment according to a computer-generated randomisation list."
Allocation concealment (selection bias)	Low risk	Personal communication with first author: "The details of the randomisation series were unknown to any of the investigators and were contained in a set of sealed opaque envelopes."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	An economic evaluation was conducted alongside a double-blind,multinational, randomised clinical trial. Personal communication with first author: "This means that both investigator and patient were blinded regarding allocation to treatment."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	"Data at baseline consisted of self-reported patient questionnaires recording use of healthcare services and days of sick leave .. .." Personal communication with first author: "Patients were blinded regarding allocation to treatment."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	"Depressive symptoms were assessed by trained raters." Personal communication with first author: "Outcome assessors were blinded for the allocation of patients."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow-up depression data is 15%, which we consider high and no appropriate method has been used to account for attrition "Efficacy analyses were conducted on the intention-to-treat (ITT) population, which included all randomised patients who took at least 1 dose of double-blind study medication and who had at least 1 valid post-baseline assessment of the MADRS total score. The ITT population thus comprised 146 patients in the escitalopram group and 142 patients in the venlafaxine group. A total of 249 patients (of 293) completed the study."

Fernandez 2005 (*Continued*)

Incomplete outcome data (attrition bias) Sick Leave	High risk	Lost to follow-up sick leave data is 16%, which we consider high and no appropriate method has been used to account for attrition "Data at baseline consisted of self-reported patient questionnaires recording use of healthcare services and days of sick leave Of the 293 patients in the trial, valid cost information in the 3-month pre-study period was available for 251 patients; for 22 patients in the escitalopram arm and 20 patients in the venlafaxine arm, either the physician or patient did not fill in the resource use questionnaire. Of the 251 evaluable patients, 126 received escitalopram and 125 received venlafaxine. Of these, 245 patients reported valid cost information for the 8-week duration of the trial (four escitalopram and two venlafaxine patients were lost relative to the pre-study period) "Given the very low rate of attrition in the sample during the trial, patients with missing data were unlikely to represent serious bias to the results of the present analysis. As a result, no attempt was made to impute missing data."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None Identified

Hees 2013

Methods	Two armed RCT. Recruitment: Between December 2007 and October 2009, participants were referred by occupational physicians from several occupational health services. Follow up: 18 months. Lost to follow up: 13.7%
Participants	117 were randomised (T1: 39; T2: 78); Setting: Outpatient; Department of Psychiatry, Academic Medical Center, Amsterdam; Inclusion: Age 18 to 65, DSM-IV diagnosis of MDD, Absent from work at least 25% of their contract hours due to their depression. In addition, the duration of the depression had to be at least 3 months or the duration of their sickness absence had to be at least 8 weeks. Finally, there had to be a relation between the depressive disorder en the work situation, that is, work was one of the determinants of depressive disorder and contributed substantially (> 25%), or the depressive symptoms reduced productivity or hindered RTW Exclusion: severe alcohol or drug dependence, bipolar disorder, psychotic disorder, depression with psychotic characteristics, indication of inpatient treatment

	Age: T1: 41.5 (SD 9.6); T2: 43.8 (SD 9.0) Male: T1: 41%; T2: 53% Education (years): T1: 13.9 (SD 3.7); T2: 13.5 (SD 3.1) Martital status: T1: 59% married or living together; T2: 58% married or living together; T1: 23% single; T2: 28% single; T1: 18% divorced or widowed; T2: 14% divorced or widowed Contract (number of hours): T1: 32.7 (SD 5.8); T2: 35.0 (SD 5.0) Absenteeism (number of hours): T1: 27.1 (SD 8.8); T2: 27.6 (SD 10.0) Duration of absenteeism (months): T1: 3.8 (IQR 2.0 - 6.5); T2: 5.0 (IQR 2.8 - 5.0) Occupational sector: financial or insurance: T1: 54%; T2: 58%; Health care: T1: 18%; T2: 9%; Other: T1: 28%; T2: 33% Work experience (years): T1: 14.1 (SD 9.6); T2: 15.9 (SD 11.0)
Interventions	T1: Treatment as usual: treatment by psychiatric residents in an outpatient university clinic according to a treatment protocol consistent with the APA guidelines. 19 visits consisted of clinical management, including psycho education, supportive therapy and cognitive behavioural interventions. Therapies were supervised on a weekly basis by an experienced senior psychiatrist specialised in depression. If needed, participants received pharmacotherapy according to a protocolised algorithm. If the participant's condition deteriorated and outpatient treatment was no longer deemed adequate, he or she was referred to day treatment or inpatient treatment T2: Adjuvant occupational therapy: consisted of 18 sessions (nine individual sessions, eight group sessions and a meeting with the employer), and was conducted by two experienced occupational therapists who had received extensive training in the intervention protocol. During the intervention, the occupational therapist frequently communicated with the occupational physician and the resident treating psychiatric. Employees were recruited to work at least 2 hours per week when starting OT, so that employees were able to directly practise the things learned (e.g. new coping strategies) during therapy
Outcomes	Absenteeism: 1) work participation, defined in: a) average number of hours of absenteeism over each 6-month period and b) duration of sick leave due to depression in calender days from the start of treatment until partial (or full) RTW. Time until partial or full RTW was operationalised as the duration of sick leave due to depression in calendar days from the start of treatment until partial (or full) RTW. Partial RTW was defined as working an increment of at least 5 hours (compared with hours worked at baseline), for at least 4 weeks without partial or full recurrence. Full RTW was defined as working the full number of contract hours in own or other work for at least 4 weeks, without partial or full recurrence Clinical: 1) severity of depression, assessed by the Hamilton Rating Scale for depression (HRSD) 2) depression remission, defined as having HRSD ≤ 7 3) severity of depression, assessed by the Questionnaire Inventory of Depressive Symptoms Self-Report (StIDS-SR) Functioning: 1) at work functioning: weekly self-report records of work efficiency on a scale 1-0 and 3 sub scales of WLQ: Output, time, mental-interpersonal 2) health-related functioning, 3 subscales of MOS-SF 36: role limitations due to emotional problems, mental health, role limitations due to physical problems

Notes		
Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Randomization was conducted by an independent research assistant, using software based on a minimization randomization procedure."
Allocation concealment (selection bias)	Low risk	"Randomization was conducted by an independent research assistant, using software based on a minimization randomization procedure."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	"Due to the nature of the intervention, neither patients nor therapists could be blinded to the patient's allocation status. " Both treatments cannot be considered equally desirable for patients, so risk of performance bias high
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sickness absence data are measured by the use of self-report. As patients are not blinded for the allocation status, risk of bias is high
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	"Study assessment were conducted by a psychiatrist and a researcher who where blind to group allocation." As the HRSD is a clinician-rated instrument, there is a low risk of bias for the HRSD outcome
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Lost to follow up: T1: 15.4%; T2: 12.8% but appropriate imputation methods have been used. "To take potential biased outcomes caused by selective loss to follow up into account, we used multiple imputation (five imputed datasets), which, assuming missing at random for missing values, gives unbiased results with correct SEs."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Lost to follow up: T1: 15.4%; T2: 12.8% but appropriate imputation methods have been used. "To take potential biased outcomes caused by selective loss to follow up into account, we used multiple imputation

		(five imputed datasets), which, assuming missing at random for missing values, gives unbiased results with correct SEs."
Selective reporting (reporting bias)	Low risk	The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way
Other bias	Unclear risk	None identified

Hollinghurst 2010

Methods	RCT. Recruitment: patients were recruited from 55 general practices in Bristol, London, and Warwickshire between October 2005 and February 2008. Follow up: 8 months. Lost to follow up: 53% for sickness absence and 29% for clinical outcomes
Participants	297 were randomised (T1: 149; T2: 148). Setting: patients between who where identified in primary care as having a new episode of depression Inclusion: patients between 18 and 75 who where identified in primary care as having a new episode of depression which was defined as being diagnosed within the 4 weeks preceding referral. Depression was defined as a score of 14 or more on the BDI12 and an ICD-10 diagnosis of depression using the CIS-R) Exclusion: patients treated for depression in the 3 months before the present episode, patients with a history of bipolar disorder, psychotic disorder, alcohol or substance misuse, and those already receiving psychotherapy Female: T1: 69%; T2: 67% Age: T1: 35.6 (SD 11.9); T2: 34.4 (SD 11.3) Marital status: T1: 34% married; T2: 39% married T1: 50% single; T2: 47% single T1: 16% separated or divorced or widowed; T2: 15% separated or divorced or widowed Employment status: T1: 65% employed; T2: 56% employed T1: 15% student; T2: 24% student T1: 20% not in employment; T2: 20% not in employment Highest educational level: T1: 65% A level or above; T2: 63% A level or above T1: 32% other; T2: 33% other T1: 3% no educational qualifications; T2: 4% no educational qualifications
Interventions	T1: Online CBT in addition to usual care: participants receiving online CBT were offered up to ten sessions of 55 minutes, to be completed within 4 months from the date of randomisation when possible. Each participant was assigned their own therapist for the duration of the study. Participants and therapists typed free text into the computer, with messages sent instantaneously, using only this means of communication T2: Usual care from GP while on a 8-month waiting list for online CBT: participants on the waiting list were not to receive psychotherapy during the study follow-up period.

	Those on the waiting list who had still an eligible Beck Depression Inventory (BDI) score after 8 months were offered the intervention at that time
Outcomes	Absenteeism: 1) the number of working days lost because of depression (time off work) over 8 months Clinical: 1) depression severity, assessed by the BDI 2) recovery, defined as a score of less than 10 on the BDI
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Randomization was by means of a computer-generated code, implemented by an individual who was not involved in the recruitment process, and communicated to the participant within 48 h of the baseline interview."
Allocation concealment (selection bias)	Low risk	"Randomization was by means of a computer-generated code, implemented by an individual who was not involved in the recruitment process, and communicated to the participant within 48 h of the baseline interview." "The allocation was concealed in advance from participants, researchers involved in recruitment, and therapists."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Risk of performance bias is considered high as patients were aware of their allocation status and both treatments are not equally desirable for patients: "Randomization was by means of a computer-generated code, implemented by an individual who was not involved in the recruitment process, and communicated to the participant within 48 h of the baseline interview."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	The number of working days lost because of depression was recorded in a diary by the participants themselves. As participants were aware of their intervention status, risk of bias high
Blinding of outcome assessment (detection bias)	High risk	The BDI is a self-report inventory. As participants were aware of their intervention

Hollinghurst 2010 (*Continued*)

Depressive symptoms		status, risk of bias high
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up is high: T1: 27%; T2: 32% even though appropriate method has been used to account for these missing data: "Fourth, a sensitivity analysis investigated the effect of missing data with multiple imputation by chained equation methods in Stata." "Analyses imputing missing values suggested that differences in attrition between the groups did not introduce any noticeable bias."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up is high: T1: 50%; T2: 55% even though appropriate method has been used to account for this missing data: "we imputed missing observations of cost and QALYs using the multiple imputation by chained equation procedure in Stata release 10." "We acknowledge that more complete data would have been available if we had used questionnaires completed face to face or data from practice records. However, the results of the imputation suggest that any information lost is unlikely to have a major influence on the results or conclusions."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Kendrick 2005

Methods	RCT, randomisation on the level of patients stratified for referring GP; 3 conditions. Recruitment: general practices referred patients to the study. CMHNs were employed by local NHS trusts. Follow up: 26 weeks. Lost to follow up: 26%
Participants	247 randomised (T1: 90; T2: 79; T3: 78). Setting: community mental health, UK. Inclusion: age: 18-65; new episode of anxiety, depression or reaction to life difficulties; minimum duration symptoms: 4 weeks; maximum duration symptoms: 6 months; GHQ-12 score at least 3 Exclusion: patient already in contact with psychiatric services; Patient already receiving psychological treatment; Severe mental illness such as schizophrenia, manic-depressive psychosis; severe substance misuse, dementia or severe depression with active suicidal ideas; housebound patients; patients without the spoken and written language skills necessary to participate; seriously ill and terminally ill patients; temporary residents Mean age: T1: 35.8 (SD 10.92); T2: 34.2 (SD 11.33); T3: 34.9 (SD 11.77)

	Female: T1: 72%; T2: 70%;T3: 69% Married or cohabiting: T1: 60%; T2: 58%; T3: 48% Fulltime or part-time employed: T1: 66%; T2: 75%; T3: 69%
Interventions	T1: CMHN problem-solving treatment: 1. explanation of treatment and rationale 2. clarification and definition of problems 3. choice of achievable goals 4. generations of alternative solutions 5. selection of preferred solution 6. clarification of necessary steps to implement solution 7. evaluation of progress; Initial 1-hour session + 5 follow-up sessions of 30-45 minutes. T2: Generic CMHN; nurses were asked to use whatever treatment they were experienced in giving; initial 1-hour session + 5 follow-up sessions of 30 to 45 minutes. Range 0 to 8 sessions T3: GP care: usual care, but asked not to refer patients to a psychological therapist during the study period unless absolutely necessary
Outcomes	Absenteeism: 1) number of days off paid work Clinical: 1) CIS-R 2) HADS-D Productivity: 1) SAS, however, subscale "work outside the home" not separately reported
Notes	Personal communication: data for depressed subsample was provided

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"The telephone randomisation service at the university of York was contracted."
Allocation concealment (selection bias)	Low risk	"Remote central randomisation was provided by telephone" "Randomisation sequences were in block sizes of either three or six, to prevent practitioners from guessing to which arm the next referral would be."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	High risk for the comparison with the GP usual care group (T3) as this treatment cannot be considered equally desirable as T1 and T1 for patients and patients were not blinded. "Table 16: n = 50 received their preferred treatment; n = 114 did not receive their preferred treatment; n = 83 reported

		no preference"
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "Number of days off paid work was captured by a resource-use questionnaire filled out by patients." "Patients were reminded not to reveal their allocation at the follow-up assessments."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression symptom score (CIS-R and HADS-D) were measured by self-report and patients were not blinded. "The computerised version of the CIS-R, which is self-complete, was used in this study." "Patients were reminded not to reveal their allocation at the follow-up assessments."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up is considered to be high (26%). Risk of attrition bias due to follow-up losses is therefore considered to be high, although sensitivity analyses were conducted and the authors conclude otherwise; "sensitivity analyses were conducted to see whether the result changed depending on what assumptions were made about the missing data". "Table 12 shows that the main findings are not particularly sensitive to the different assumptions about missing data that were investigated." It was harder to retain patients in the GP care (thus higher loss to follow up in that group): "Although the overall follow-up rates were good, there was a lower follow-up rate in the GP arm. It is difficult to tell whether this biased the findings in a particular direction. Follow-up rates were better among those patients who received the treatment they preferred, so it is likely that there were more disaffected patients in the GP care arm. However, it is not known whether those who dropped out remained more symptomatic than those who were followed up. Failing to receive their treatment of preference was not associated with a worse outcome on the CIS-R among those who were followed up. The sensitivity

		analyses suggest that CMHN care, whether generic care or specific PST, is unlikely to be more effective than GP care, unless one believes the LOCF analysis and makes the extreme assumption that all the dropouts remained as symptomatic as they were at the time of last assessment."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up for sick leave data is considered to be high (26%). Risk of attrition bias due to follow-up losses is therefore considered to be high, although sensitivity analyses were conducted and the authors conclude otherwise; "cost results from this analysis were validated by substituting where possible data from the GP case notes in place of imputed values for missing data, and repeating the analysis. Overall, the results did not change significantly. " "36% had at least one resource item missing over the 6-month follow up. Therefore, complete resource use data were available for 159 (64%) of the patients. The results presented here are based mainly on the 184 patients for whom complete CIS-R data were available over the 6-month period. To achieve this sample, 25 (14%) of the patients who had CIS-R data but not resource-use information had to be imputed. The results were then compared with those obtained using data from GP notes where available instead of imputation, and those obtained using only the 159 patients with complete resource-use data. After imputing missing values for the 25 patients with missing resource-use data, the numbers of patients included in the economic analysis in each group were as follows: 51 patients in GP care (28%), 62 patients in generic CMHN care (34%) and 71 patients in PS CMHN care (38%)."
Selective reporting (reporting bias)	Low risk	No indication for selective reporting could be identified. However, in the design study, the comparisons of T1 with T2 was not pre-specified
Other bias	Unclear risk	None identified

Knekt 2013

Methods	RCT. Recruitment: a total of 459 eligible outpatients were referred to the Helsinki Psychotherapy Study from psychiatric services in the Helsinki region from June 1994 to June 2000. Follow up: 5 years. Lost to follow up: 19% (for all participants over five years), lost to follow up for the subgroup of people with depressive disorder: 51% (over five years)
Participants	326 were randomised (T: 97; T2: 101; T3: 128). Subgroup of people with depressive disorder: 161. Setting: outpatient. Inclusion: 20 to 45 years of age and suffered from a longstanding (> 1 year) disorder causing dysfunction in work ability. They were also required to meet DSM-IV criteria for anxiety or mood disorders Exclusion: psychotic disorder or severe personality disorder, adjustment disorder, substance-related disorder, organic brain disease or other diagnosed severe organic disease, and mental retardation. Individuals treated with psychotherapy within the previous 2 years and psychiatric health employees were also excluded Age: T1: 33.6 (SD 7.2); T2: 32.1 (SD 7.0); T3: 31.6 (SD 6.6) Male: T1: 25.8%; T2: 25.7%; T3: 21.1% Employed or student: T1: 83.2%; T2: 85.1%; T3: 75.4% Academic education: T: 28.9%; T2: 19.8%; T3: 75.4%
Interventions	T1: Solution-focused therapy: is a brief, focal, transference-based therapeutic approach which helps patients by exploring and working through specific intrapsychic and interpersonal conflicts. The therapy included one session every second or third week, with a limit of 12 sessions, over no more than 8 months T2: Short-term psychodynamic psychotherapy: is characterized by the exploration of a focus, which can be identified by both the therapist and the patient. This consists of material from current and past interpersonal and intrapsychic conflicts and the application of confrontation, clarification, and interpretation in a process in which the therapist is active in creating the alliance and ensuring the time-limited focus. The therapy was scheduled for 20 weekly treatment sessions over 5 to 6 months T3: Long-term psychodynamic psychotherapy: is an open-ended, intensive, transference-based therapeutic approach which helps patients by exploring and working through a broad area of intrapsychic and interpersonal conflicts. The therapy is characterized by a framework in which the central elements are exploration of unconscious conflicts, developmental deficits, and distortions of intrapsychic structures. Confrontation, clarification and interpretation are major elements, as well as the therapist's actions in ensuring alliance and working through the therapeutic relationship to attain conflict resolution and greater self-awareness. Therapy includes both expressive and supportive elements, the use of which depends on patient needs. The frequency of sessions was 2 to 3 times a week, and the duration of the therapy was up to 3 years
Outcomes	Absenteeism: 1) number of sick-leave days during last 3 months Clinical: 1) depressive symptoms assessed by the Beck Depression Inventory (BDI) 2) depressive symptoms assessed by the Hamilton Depression Rating Scale (HDRS) Functioning: 1) the work-subscale (SAS-work) of the social adjustment scale (SAS-SR)
Notes	

Knekt 2013 (*Continued*)

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Concealed assignment codes were given sequentially to patients in consecutively numbered envelopes."
Allocation concealment (selection bias)	Low risk	"The patients who fulfilled the selection criteria at baseline were randomized into solution-focused therapy, short-term psychodynamic psychotherapy or long-term psychodynamic psychotherapy or long-term psychodynamic psychotherapy in a 1:1:1.3 ratio using a central computerized randomization schedule. Concealed assignment codes were given sequentially to patients in consecutively numbered envelopes."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Due to the nature of the intervention, the participants and personnel could not be blinded
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and the patients were not blinded for their allocation status. Outcome is likely to be influenced by this lack of blinding. "The number of sick leave days from work during the past 3 months were collected by single-item questions included in a follow-up questionnaire developed in the project." "Unavoidable weaknesses in a study like this are [...] the lack of blindness of assessments."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	The BDI is a self-report inventory and patient were not blinded for their allocation status. Outcome is likely to be influenced by this lack of blinding. The HDRS is a clinician-administered scale but clinicians were also not blinded: "raters were not blinded since they were provided with information on the treatment group at the five interview sessions during the 3-year follow up."
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Loss to follow up is 19% and missing values were replaced by multiple imputation; this did not alter the results. "Analyses based

		on multiple imputation and taking into account the need for treatment at the time of dropout did not, however, notably alter the results, suggesting that the results presented are unbiased (data not shown)."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up is considered to be high: 39% at one year and 52% at five years
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Krogh 2009

Methods	Randomized pragmatic trial. Recruitment: between January 2005 and July 2006. Follow-up: 12 months. Lost to follow up: 17% at 4 months and 22% at 12 months
Participants	165 were randomised (T1:55; T2:55; T3:55); Setting: outpatient; this trial was carried out at a single location at Copenhagen University. Inclusion: age 18-55 years, referred by a medical doctor or psychologist, meeting ICD-10 criteria for unipolar depression, living in the Greater Copenhagen catchment area, able to read and understand informed consent. Exclusion: being engaged in regular sports activity for more than 1 hour per week, ongoing alcohol or substance abuse judged to be at risk of suicide, poor Danish language skills, having a medical condition that contraindicated physical exercise, or had been on sickness leave for than 24 consecutive months Age: T1: 41.9 (SD 8.7); T2: 38.1 (SD 9.0); T3: 36.7 (SD 8.7) Female: T1: 81.8%; T2: 78.2%; T3: 61.8% Ethnicity: T1: 90.9% Caucasian; T2: 92.7% Caucasian; T3: 90.9% Caucasian Occupational status: T1: 41.8% unemployed; 40% fulltime work; 14.5% part-time work; 3.6% < 20 hrs/wk T2: 54.5% unemployed; 32.7% fulltime work; 10.9% part-time work; 1.8% < 20 hrs/wk T3: 36.4% unemployed; 41.8% fulltime work; 18.2% part-time work; 3.6% > 20 hrs/wk
Interventions	T1: Supervised strength training. Designed to increase muscular strength, initially with 12 repetitions of 50% of repetition maximum 2 or 3 times per exercise. As the patients progressed, the numbers of repetitions were reduced to 10 and 8, with an increase of RM to 75%. The training was a circuit-training program with 6 exercise on machines involving large muscle groups. As a supplement to this, free weights and sandbags were used for exercising the calf muscles, the arm abductors, the triceps muscles, and the hip abductors. All patients were scheduled to meet twice per week during a 4-month period for a total of 32 sessions T2: Aerobic training. Designed to increase fitness as measured by maximal oxygen uptake. The program involved 10 different aerobic exercises using large muscle groups. Machines were used for cycling, running, stepping, abdominal exercises, and rowing. Additional exercises were sliding movements on small carpets, trampoline, step bench, jump rope,

	and Ski Fitter. During the first 8 sessions, each exercise was done twice for 2 minutes with a 2-minute rest at an intensity level of 70% of maximal heart rate. This gradually increased to a level at which exercise was done for 3 minutes with a 1-minute rest at an intensity level of 89% during the last 8 sessions. All patients were scheduled to meet twice per week during a 4-month period for a total of 32 sessions T3: Relaxation training. Designed to avoid muscular contractions or stimulation of the cardiovascular system, and the patients did not engage in activities perceived higher than 12 on the Borg Scale. The first 20 to 30 minutes were used for exercises on mattresses or Bobath Balls or back massage using a Ball Stick Ball. This was followed by light balance exercises for 10 to 20 minutes and by relaxation exercises with alternating muscle contraction and relaxation in different muscle groups while lying down for 20 to 30 minutes
Outcomes	Absenteeism: 1) self-reported percentage of days absent from work during the last 10 working days at 4 and 12 months Clinical: 1) severity of depression, assessed by the Hamilton Rating Scale for Depression (HAM-D17) 2) remission, defined as not fulfilling the ICD-10 criteria for depression and having a HAM-D17 < 8 3) severity of depression, assessed by the Beck Depression Inventory (BDI) Employment status: 1) % on sick leave 2) % unemployed
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Randomization was carried out by the CTU using computerized restricted randomization with a block size of 6. The block size and thus the allocation sequence were unknown to the DEMO trial staff." "The strengths of our trial were the centralized randomization, which provided adequate generation of the allocation sequence and adequate allocation concealment"
Allocation concealment (selection bias)	Low risk	"Randomization was centralized and stratified according to medicine status." "The strengths of our trial were the centralized randomization, which provided adequate generation of the allocation sequence and adequate allocation concealment"

Blinding of participants and personnel (performance bias) Sick Leave	High risk	"The same 2 physiotherapists were used throughout the trial period. The type and number of exercise interventions were distributed evenly between the two, and thus the physiotherapists were not blinded to allocation". "And the patients were instructed not to reveal their group assignment." "The lack of blinding of treatment allocation for patients and psychotherapists could lead to collateral interventions, possibly confounding our results." As the relaxation condition was not equally desirable to patients as the other two groups, the risk of performance bias is considered high
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Absenteeism measured by self-report. As patients were aware of their allocation status, risk of bias high
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	For HAM-D17: "The assessor was blinded to intervention group, and the patients were instructed not to reveal their group assignment. After assessment the assessor was requested to guess which group the patient has been assigned to, making it possible to examine whether the blinding was successful [..] This indicated that the blinding of the assessors was successful"
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up at endpoint was high: 22% (36/165) and skewed. Risk of attrition bias was therefore considered high although an appropriate method was used to deal with missing values in the analyses and the authors conclude otherwise "Analysis of age, sex, HAM-D17, or absence from work during the last 10 working days at entry did not suggest any significant differences between missing participants and participants included in the analysis at either 4 months or 12 months." "It is then plausible to consider the missing data as 'missing at random', making the mixed effect model a plausible approach to estimate the effect, based on the total sample with missing cases included." "This approach uses data from all included patients (intention-to-treat), handles entry

		differences, and is able to handle missing data (restricted maximum likelihood procedure) with higher precision and power compared to more traditional methods such as the last observation carried forward. " "There was skewed attrition, and the follow-up assessment was significantly later than 4 months in the control group."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up at endpoint was high: 22% (36/165) and skewed. Risk of attrition bias was therefore considered high although an appropriate method was used to deal with missing values in the analyses and the authors conclude otherwise "Analysis of age, sex, HAM-D17, or absence from work during the last 10 working days at entry did not suggest any significant differences between missing participants and participants included in the analysis at either 4 months or 12 months." "It is then plausible to consider the missing data as 'missing at random', making the mixed effect model a plausible approach to estimate the effect, based on the total sample with missing cases included." "This approach uses data from all included patients (intention-to-treat), handles entry differences, and is able to handle missing data (restricted maximum likelihood procedure) with higher precision and power compared to more traditional methods such as the last observation carried forward. " "There was skewed attrition, and the follow-up assessment was significantly later than 4 months in the control group."
Selective reporting (reporting bias)	High risk	In the study protocol, no report was made regarding the third treatment group (relaxation)
Other bias	Unclear risk	None identified

Krogh 2012

Methods	A single-centre, two-armed, parallel-group, observer-blinded randomised clinical superiority trial. Recruitment: between September 2008 and April 2011, participants were referred to trial site from various clinical settings. Follow up: 3 months. Lost to follow up: 13%
Participants	115 were randomised (T1: 56; T2: 59). Setting: outpatient; the participants were enrolled at the trial site in Copenhagen (Denmark) from various clinical settings. Inclusion: men and women between 18 and 60 years of age, referred from a clinical setting by a physician or a psychologist, a diagnose of major depression (DSM-IV) based on the Danish version of the Mini International Neuropsychiatric Interview, score above 12 on the HAM-D17 and living in the Greater Copenhagen catchments area, able to comprehend and sign the informed consent statement Exclusion: current drugs abuse, any antidepressant medication within the last two months, current psychotherapeutic treatment, contraindications to physical exercise, more than 1 hour or recreational exercise per week, suicidal behaviour according to the 17-item Hamilton depression rating scale (HAM-D17 item 3 > 2), pregnancy, current/previous psychotic or manic symptoms, or lack of informed consent Age: T1: 39.7 (SD11.3); T2: 43.4 (SD 11.2) Female: T1: 71.4%; T2: 62.7% Occupational status: T1: 35.7% unemployed; T2: 45.7% T1: 35.7% sickness leave; T2: 30.5% sickness leave T1: 74.3% job attendance, last 10 days; T2: 73.8% job attendance, last 10 days
Interventions	T1: Aerobic training group: designed to increase fitness as measured by maximal oxygen uptake. After initial 10 minutes of general low-intensity warm-up, the participants did 30 minutes of aerobic exercise on a stationary cycle ergometer followed by five minutes low-intensity cool down period. During the initial four weeks, the aim was to work out at intensity levels corresponding to at least 65% to their maximal capacity, progressing to 70% and 80% during the second and third month, respectively. The participants carried a pulse monitor during exercise to guide and document intensity levels T2: Stretching exercise group: designed as an attention control group with the purpose of providing the same level of social interaction and contact with health care professionals as in the aerobic exercise group. This was done in order to assess the potential antidepressant effect of aerobic exercise in it self, and not the effect of aerobic exercise plus social interaction. This stretching exercise group performed low intensity exercise, which we did not expect to contain any antidepressant effect per se. The initial 10 minutes were low-intensity warm-up on a stationary bike, then a 20 minutes program of stretching, followed by 15 minutes of various low intensity exercises such as throwing and catching balls Both groups were scheduled to meet three times per week for three months for a total of 36 sessions
Outcomes	Absenteeism: 1) the number of days spent on the job within the last ten working days, expressed as a percentage Clinical: 1) depression severity, assessed by the HAM-D17 2) core depression items, assessed by HAM-D6 3) remission, defined as not fulfilling the DSM-IV criteria for major depression and a

	HAM-D17 score below 8 4) self-reported depression, assessed by the Beck Depression Inventory (BDI) Employment status: 1) employment status or sick leave at the time of the interview
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"The randomization was centralized and carried out by the Copenhagen Trial Unit (CTU) using a computerized randomization sequence with alternating block sizes unknown to the investigators."
Allocation concealment (selection bias)	Low risk	"The randomization was centralized and carried out by the Copenhagen Trial Unit (CTU) using a computerized randomization sequence with alternating block sizes unknown to the investigators."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	"Prior to the first training session of the participant, the trial psychotherapist would contact the CTU by phone for participant allocation." "Neither participants nor the physiotherapist conducting the intervention were blinded to the allocation."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	The outcome assessors were all blinded to participant allocation "Prior to the follow up interview, participants were instructed not to reveal their allocation to the outcome assessors. The statistical analysis and preparation of the first draft was carried out blinded to group assignment."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	The outcome assessors were all blinded to participant allocation. The HAM-D17 is a structured interviewer based questionnaire, so risk of bias low (this does not apply to the BDI as this is a self-report instrument)
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Lost to follow up: T1: 16.1%; T2: 10.2% but appropriate method has been used to account for these missing data: "All continuous outcome measures were analyzed us-

		ing a repeated measurement linear mixed effect model with an unstructured variance matrix [..] The mixed effects function is able to handle missing continuous data using a likelihood estimation of missing data. “
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Lost to follow up: T1: 16.1%; T2: 10.2% but appropriate method has been used to account for these missing data: ” “All continuous outcome measures were analyzed using a repeated measurement linear mixed effect model with an unstructured variance matrix [..] The mixed effects function is able to handle missing continuous data using a likelihood estimation of missing data. ”
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None reported

Lerner 2012

Methods	RCT. Recruitment: 6 months. Follow up: 4 months. Lost to follow up: 8.9%
Participants	79 were randomised (T1:52; T2:27); Setting: workplace; this study involved State Government in Maine Inclusion: ages 18 to 62 years and employed 15 hours per week or more and fulfilled the criteria for current MDD and/or dysthymia, a WLQ productivity loss of at least 5% in the past 2 weeks (this score is consistent with an impaired ability to work approximately 20% of the time over 2 weeks). Exclusion: planning to retire within 2 years, receiving work disability benefits, active alcoholism or drugs-abuse based on the five-item CAGE, pregnant or 6 months postpartum, schizophrenia or bipolar disorder, non-English speaking and/or reading, and/or diagnosed with one or more of 12 medical conditions that have symptoms that potentially interfere with working (e.g. angina, congestive hart failure, stroke, diabetes, chronic obstructive lung disease) Comorbidity: T1: 80.8%; T2: 71.1% Age: T1: 45.5 (SD 9.8); T2: 45.9 (SD 8.6) Male: T1: 23.1%; T2: 18.5% Ethinicity: T1: 100% white; 96.3% white Marital status: T1: 47.1% married; T2: 48.1% married
Interventions	T1: Work and Health Initiative (WHI) intervention. Provided over the phone by EAP counsellors trained in its methods. The program lasts for 8 weeks with 1-hour visits occurring every 2 weeks. This multi component work-focused programs consists of: 1) work coaching and modification, 2) care coordination, 3) cognitive-behavioral strategies. In the WHI, the counsellor and employee co-create a care plan for dealing with each

	functional problem and review specific assignments and progress at each session. A motivational enhancement approach is utilized to promote and solidify change. In both groups: electronic feedback on depression and advise to seek care T2: Usual care. Primary care, specialty care, behavioral health programs, and/or standard EAP services. In both groups: electronic feedback on depression and advise to seek care
Outcomes	Absenteeism: 1) the WLQ Work Absence Module measured self-reported time missed from work in the past 2 weeks because of health or medical care Clinical: 1) change in depression symptom severity as measured by the PHQ-9 Functioning: 1) Work limitations Questionnaires Short Form (WLQ), a self-report survey tool for assessing the impact of health problems, including at-work performance. 4 Dimensions of performance are measured
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Employees were allocated by electronic randomization."
Allocation concealment (selection bias)	Low risk	Web-based randomisation
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Participants received information about the RCT and were thus aware of the treatment condition to which they were randomised. Seven counsellors volunteered to conduct the WHI intervention
Blinding of outcome assessment (detection bias) Sick Leave	High risk	The WLQ Work absence module is a self-report measure. As participants were aware of their allocation status, risk of bias high
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	The PHQ-9 relies on patient self-report. As participants were aware of their allocation status, risk of bias high
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	"Five (9.6%) employees in the WHI treatment group and 2 (7.4%) of the usual group did not complete the follow-up questionnaire and were considered dropouts. " "Sensitivity analyses including the seven employees that were lost to follow-up confirmed the results."

Lerner 2012 (*Continued*)

Incomplete outcome data (attrition bias) Sick Leave	Low risk	"Five (9.6%) employees in the WHI treatment group and 2 (7.4%) of the usual group did not complete the follow-up questionnaire and were considered dropouts. " "Sensitivity analyses including the seven employees that were lost to follow-up confirmed the results."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

McCrone 2004

Methods	RCT, 2 conditions. Recruitment: by screening in the GP waiting rooms and of GP referrals using the GHQ-12. Score at least 4: seen by GP who administered inclusion and exclusion criteria. Follow up: 6 months. Lost to follow up at 6 months: T1: 27%; T2: 24%
Participants	274 were randomised (T1: 146; T2: 128). Setting: Primary care, UK Inclusion: GP patients aged 18 to 75 years; diagnosis (ICD): depression, mixed anxiety/depression or anxiety disorder. CIS-R score at least 12 Exclusion: active suicidal ideas, Psychotic disorder, organic mental disorder or alcohol or drug dependence. Having taken medication for anxiety or depression continuously for at least 6 months immediately prior to entry; unable to read or write; unable to attend 8 sessions at practice Mean age: T1: 43.6 (SD 14.3); T2: 43.4 (SD 13.7) Female: T1: 73% T2: 75% Married or cohabiting: T1: 54%; T2: 52% Employed: T1: 66%; T2: 58%
Interventions	T1: Computerised CBT: interactive, multimedia. Feedback to patient and GP after each session. 15 minute introductory video, 8 x 50 minute sessions of CBT, with homework projects between sessions T2: TAU: General practitioner care as usual: no constraints. Could include medication, discussion of problems with GP, practical or social help, referral to counsellor, practice nurse, mental health professional, or further physical examination
Outcomes	Absenteeism: 1) Number of days of absence from work (certified by GP) during 8 months Clinical: 1) BDI Productivity: 1) Work and Social Adjustment Scale
Notes	

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"The random allocation schedule was generated at the Institute of Psychiatry. An individual unit of randomization was used."
Allocation concealment (selection bias)	Low risk	"Random allocation schedule was generated at the Institute of Psychiatry, before the study commenced and away from GP practices. Cards in sealed and numbered envelopes were used. Only to be opened by practice nurse who ran study. Integrity was checked by the first author on her regular visits to the practices."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	No blinding, risk of performance bias considered high as the treatment of interest (T1) cannot be considered equally desirable as Treatment as usual (T2) for patients. "Patients randomized to 'Beating the Blues' (T1) also received pharmacotherapy, if prescribed by their GP, and/or general GP support and practical/social help", offered as part of treatment as usual, with the exception of any face-to-face counselling or psychological intervention. We did not constrain the interventions received by patients allocated to treatment as usual (T2) ." Moreover, patients in the Treatment as Usual (T2) group were found to attend other health care professionals more often. "Large differences were observed for the proportion of patients attending accident and emergency or outpatient departments, and having contacts with community psychiatric nurses, counsellors and other therapists. Greater use was made by the TAU group for all these services."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	No blinding of outcome assessors was reported. Sick leave was based on the sick leave certificates of the GP, who was also the treatment provider of treatment as usual. "We recorded the number of days of absence from work during the baseline and follow-up periods on the basis of an issue

McCrone 2004 (*Continued*)

		of a certificate by the general practitioner."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	No blinding of patients was reported and depressive symptoms were measured by self-report "Depressive symptoms were measured with self-report and participants were not blinded to treatment allocation."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up was relatively high (> 20%) for the depression outcome From Figure 2 of the publication on depression outcome (Proudfoot et al 2004): Loss to follow up: T1: 27%; T2: 24%
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Sick leave data were part of the cost data, and a high percentage of the cost data were complete at follow up. "A total of 274 patients were randomised into two groups (BtB, n = 146; TAU, n = 128), with cost data available for both baseline and follow-up periods for 261 (95%) patients (138 BtB, 123 TAU)."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Miller 1998

Methods	RCT. multicentre, 2 conditions. Recruitment: referrals from physicians or mental health professionals, media advertising, and word of mouth. Follow up: 12 weeks. Lost to follow up: 2%
Participants	635 were randomised: (T1: 426; T2: 209). Setting: 12 outpatient centres in USA Inclusion: age 21 to 65 years; Diagnosis of chronic MDD with two or less cumulative depression-free months and who had not met DSM-II-R criteria for dysthymia within 2 months of the onset of current MD episode OR of concurrent MD episode superimposed on antecedent DSM-III-R dysthymia; Premenopausal women: adequate contraception Exclusion: organic mental syndrome, current or lifetime diagnosis of bipolar disorder or cyclothymia, schizophrenia, other psychotic disorders, obsessive-compulsive disorder, antisocial, schizotypical or severe borderline personality disorder; Principal DSM-III-R diagnosis of panic disorder, generalized anxiety disorder or PTSD within the past 6 months; DSM-II-R defined anorexia or bulimia nervosa within the past year; Drug or alcohol abuse or dependence within the past 6 months; Patients deemed at immediate suicide risk/ medical contraindications to antidepressants; Significant general medical disorder;Concomitant therapy with any psychotropic drug (except chloral hydrate or temazepam); Failure of adequate trial of sertraline or imipramine; Treatment with MOA-

	inhibitors within 3 weeks; Any depot neuroleptic within 6 months'; Fluoxetine within 1 month; Regular daily neuroleptic, anxiolytic, or antidepressant medication within 2 weeks; ECT within 3 months Mean age: 41.1 (SD 10.1) Female: 63% Married: 38% Employed: 71%
Interventions	T1: sertraline (SSRI). Week 1-3: 50 mg/day, then weekly titration in 50 mg/day increments (max 200 mg/day). 12 weeks, visits every week for the first 6 weeks and every 2 weeks for last 6 weeks. Before this, 1 week placebo run-in T2: Imipramine (TCA). Week 1: 50 mg/day, week 2: 100 mg/day, week 3: 150 mg/day. Then weekly titration 50 mg/day increments with a max of 300 mg/day by week 6. 12 weeks, visits every week for the first 6 weeks and every 2 weeks for last 6 weeks. Before this, 1 week placebo run-in
Outcomes	Absenteeism: 1) hours worked per week (12 weeks) Clinical: 1) full remission, both CGI-I (=sub scale CGI) score of 1 or2 AND total HAM-D score of 7 (or less) at last visit 2) satisfactory therapeutic response, at last visit: both CGI-I (=sub scale CGI) score of 1 or 2 AND total HAM-D score of 15 or less AND HAM-D-score reduction of at least 50% since baseline AND final GSI-S (= subscale CGI) score of 3 or less 3) 24-HAM-D 4) MADRS 5) BDI Employment status: 1) employed (yes or no) Work functioning: 1) SAS work composite 2) LIFE work functioning
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Not reported
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	"A novel statistical method was employed for unblinding patients who experienced recurrence or clinically significant worsening of symptoms." "In consultation with FDA personnel, the sponsor's statistician

		monitored the ability of each investigator to guess the treatment assignment of their patients still in the study. When breaking the blind for any patient, the statistician (R.J.M.) examined the effect of unblinding on our ability to guess the treatment assignment for the remaining patients at that site. If any of these probabilities exceeded 75%, the site agreed to refer all subsequent relapsers to a third party for treatment."
Blinding of outcome assessment (detection bias) Sick Leave	Unclear risk	Sick leave was assessed by the LIFE interview. Interviewers were blind to treatment condition. "Finally, it should be noted that while blind to treatment condition, patients and interviewers were not blind to the fact that patients were receiving active medication nor were they blind to the time of assessment (baseline, week 4, endpoint)."
Blinding of outcome assessment (detection bias) Depressive symptoms	Unclear risk	Depressive symptoms were measured with the 24 HAM-D (clinician-rated). Interviewers were blind to treatment condition. "Finally, it should be noted that while blind to treatment condition, patients and interviewers were not blind to the fact that patients were receiving active medication nor were they blind to the time of assessment (baseline, week 4, endpoint)."
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	For depressive symptoms, ITT rates of remission could be calculated for 623 (of the 635) patients, which is 98%. "See Figure 1, Keller et al, 1998."
Incomplete outcome data (attrition bias) Sick Leave	Unclear risk	Completeness of sick leave data not reported. "Sample sizes [on psychosocial variables] vary due to sporadic missing data."
Selective reporting (reporting bias)	Unclear risk	No indication for selective reporting could be identified. The design was published in a paper by Rush et al, albeit concurrently with the publications on the outcome
Other bias	Unclear risk	None identified

Noordik 2013

Methods	Two-armed cluster randomised trial. Recruitment: Recruitment of workers started in November 2006 and ended in December 2007. Workers eligible according to the OP were invited to participate. Follow up: 12 months. Lost to follow up main outcome: 10. 6% for all participants and 11% for depressed subgroup
Participants	160 were randomised (T1: 75; T2: 85). Subgroup of depressed workers: 37 (T1: 18; T2: 19). Setting: Occupational healthcare. This study was conducted in the Netherlands, where most of the workers on sick leave due to CMD visit an OP. The OP offers RTW interventions to these workers according to the evidence-based (Dutch) guidelines Inclusion: workers who were on sick leave due to CMD between 2 and 8 weeks. CMD were defined as stress-related, adjustment, anxiety or depressive disorders. Stress-related disorders were classified according to the Dutch guidelines for OP (19). Anxiety, depressive, and adjustment disorders were classified by the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) Exclusion: workers with a primary somatic disorder according to the OP and those who were not able to speak Dutch Mean age: T1: 44.9 (SD 9.8); T2: 45.9 (SD 9.8) Female: T1: 75.7%; T2: 66.7% Educational level: Low: T1: 8.7%; T2: 17.9% Middle: T1: 24.6%; T2: 23.1% High: 66.7%; T2: 59.0%
Interventions	T1: Exposure based return to work intervention (RTW-E): In the RTW-E program, workers received CAU and were gradually exposed in vivo to more demanding work situations structured by a hierarchy of tasks evoking increasing levels of anxiety, stress, or anger. The RTW-E program provided workers with several homework assignments aimed at preparing, executing, and evaluating an exposure-based RTW plan T2: Care as usual (CAU): aims to help workers regain control and rebuild social and occupational contacts and activities, according to the OP practice guidelines for CMD. The OP can support this process by using recommended methods such as stress inoculation training, cognitive restructuring, graded activity, and time contingency during the RTW
Outcomes	Absenteeism: 1) the time-to-full RTW, calculated as the number of calendar days from the first day of sick leave to the first day of full RTW. Full RTW was defined as the total number of contracted working hours per week lasting ≥28 calendar days without a recurrence of sick leave Clinical: 1) symptoms of depression, assessed by the Four-Dimensional Symptom Questionnaire (4DSQ)
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement

Random sequence generation (selection bias)	Low risk	"We performed a restricted randomization with blocks of four OPs." After randomization researcher KN informed EN about the allocation of every OP and saved the randomization file." Personal communication: "The randomization followed a schedule generation by randomization software."
Allocation concealment (selection bias)	High risk	"The validity of the results of this study may have been limited due to a selection bias because of the absence of allocation for each OP. As a result, the potential for the selective inclusion of workers was rather high." "However, we could not prevent some OP from including zero workers, which could have introduced selection bias."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Blinding of participants and researchers, but not of personnel was ensured: "The workers were blind to the differences in RTW-E and CAU." "The researchers were blind to the allocation and outcome measurement."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Sick leave was assessed by workers' diaries. As workers are blinded to allocation status, risk of detection bias for sick leave is considered to be low
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression is assessed by the 4DSQ, a self-report questionnaire. As the participants were blinded to allocation status, risk of detection bias for depressive symptoms is considered to be low
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow-up for depression for the subgroup of depressed workers: 52%
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up of sick leave data for the subgroup of depressed workers was: 11%. No appropriate method was used to take selective attrition into account
Selective reporting (reporting bias)	High risk	Not all (secondary) outcomes measures announced in the design paper were reported in the effect study, of which the data on the HADS-depression subscale
Other bias	Unclear risk	None identified

Romeo 2004

Methods	RCT. multicenter, 2 conditions. Recruitment: from general practitioners' practices. Follow up: 24 weeks. Lost to follow up: T1: 6%; T2: 14%
Participants	177 were randomised: (T1:93; T2:84). Setting: primary care, outpatients in Scotland, UK. Inclusion: > 18 years old; Depressive episode according to DSM-IV checklist; 17-HAM-D score > 18 Exclusion: schizophrenia, Bipolar, suicidal, illicit drug abuse or alcohol dependence; Treatment with any other psychotropic drug within 1 week before entry, or mirtazapine or paroxetine during the present episode, or treatment within 5 weeks before entry with fluoxetine, or any other antidepressant within 2 weeks before entry; renal, hepatic, respiratory, cardiovascular, or cerebrovascular disease; pregnancy or lactating, or no contraception Age: T1: 40 (SD 14.3); T2: 40 (SD 11.7) Female: T1: 75%; T2: 71% Fulltime or part-time employed: T1: 48%; T2: 58%
Interventions	T1: Mirtazapine (TCA): 30 tto 45 mg/day oral Week 1 - 4 30 mg/day Week 5 - 24: optional increase to 45 mg/day (discretion of the investigator) T2: Paroxetine (SSRI): 20-30 mg/day oral Week 1 - 4: 20 mg/day Week 5 - 26 optional increase to 30 mg/day (discretion of the investigator)
Outcomes	Absenteeism: 1) total mean days lost due to illness in 24 weeks Clinical: 1) primary: change from baseline on 17-HAM-D; Secondary: 17-HAM-D responder rates (= at least 50% change from baseline to endpoint); 17 HAM-D remitter rates (= % with score of 8 or less on two assessments after the first score of 8 or less)
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	A randomisation list was used that was prepared in advance "Randomization was performed according to centrally prepared randomization lists."
Allocation concealment (selection bias)	Low risk	"Randomization was performed according to centrally prepared randomization lists." Personal communication: "The person assessing eligibility for inclusion was blind to allocation concealment."
Blinding of participants and personnel (performance bias)	Low risk	Double-blind study design. Personal communication: "Medication was dispensed by

Romeo 2004 (*Continued*)

Sick Leave		the GP who was blinded to treatment allocation."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Double-blind study design. Sick leave was assessed by questionnaires filled out by patients, who were blinded to treatment allocation
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	Double-blind study design. Personal communication: "Outcomes were assessed by trained research nurses who were blind to treatment allocation."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Lost to follow-up: T1: 6%; T2: 14% and no appropriate imputation methods have been used "Six excluded mirtazapine patients, four were lost to follow-up, one dropped out early, and one refused participation in the study. Of the 14 excluded paroxetine patients, five were lost to follow-up, four were early drop outs, two did not participate any further, one discontinued due to the lack of efficacy, one was hospitalized as a results of a concomitant disease and one did not fulfil the selection criteria." "The high attrition rate observed in our study should be taken in to account when interpreting efficacy results due to possible influence on overall efficacy results."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Lost to follow-up: T1: 6%; T2: 14% and no appropriate imputation methods have been used "Six excluded mirtazapine patients, four were lost to follow-up, one dropped out early, and one refused participation in the study. Of the 14 excluded paroxetine patients, five were lost to follow-up, four were early drop outs, two did not participate any further, one discontinued due to the lack of efficacy, one was hospitalized as a results of a concomitant disease and one did not fulfil the selection criteria." "The high attrition rate observed in our study should be taken in to account when interpreting efficacy results due to possible influence on overall efficacy results."

Romeo 2004 (*Continued*)

Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Rost 2004

Methods	RCT, randomisation on the level of practice, 12 practices were randomised. Recruitment: Trained administrative staff recruited patients who made routine-length visits to physicians. They asked eligible (see inclusion) patients to participate in 2 min first stage depression screener. Patients who screened positive and did not meet exclusion criteria were immediately invited to complete 5 min second stage screener. If they screened positive, they were asked to participate in study. Follow up: 24 months. Lost to follow up: 27%
Participants	326 employed persons were randomised: (T1: 158; T2: 168). Setting: Community primary care practices across the US. Inclusion: Age > 18, sufficient literacy in English and cognitive function to complete surveys requiring 6-months recall, acces to telephone; Positive first screen: 2 weeks or more depressed or loss of interest in past year AND 1 week or more of this in last month; Second screen: 5 or more of 9 criteria for major depression in past 2 weeks on Inventory to diagnose depression. Exclusion: pregnant, breastfeeding or <3 months postpartum; Acute life-threatening physical conditions; Pos screeners who reported that symptoms started after loss of a loved one; pos screeners who did not intend to receive ongoing care in the clinic in the next year; Second stage screener: self-report lifetime mania, use of lithium or current alcohol dependence Age: T1: 37.9 (SD 10.9); T2: 40.2 (SD 10.3) Female: T1: 84.2%; T2: 85.7% Married: T1: 45%; T2: 51% Employed: 100%
Interventions	T1: Enhanced care. Primary care team was trained to provide high quality depression treatment. After enrolment, patients were evaluated for depression by physician and asked to return within one week to nurse care manager. Subsequent visit: education about treatment, addressing treatment barriers, checklist for physician's review, scheduling of next appointment in one week. This continued for 5-7 weeks. Then patients were monitored (symptoms and treatment adherence) for one year. Physicians reviewed patients monthly based on report of nurses to see whether guideline recommendations were followed. Medication algorithm of guideline: initially SSRI or secondary amine tricyclic. Switch drug classes when response failure T2: Usual Care. Regular Primary physicians care
Outcomes	Absenteeism: 1) total number of work hours lost due to illness or doctor visits over past 4 weeks Clinical: 1) depression severity: CES-D (adapted) Productivity:

	1) subjective rating on 0 to 10 scale of productivity at work	
Notes		
Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"The practices were stratified and matched into six pairs." "Within each pair, one practice was randomized to the 'enhanced' care condition and the other practice delivered usual care to study participants."
Allocation concealment (selection bias)	High risk	Personal communication: "The allocation of the practice was known to the administrative staff who screened patients."
Blinding of participants and personnel (performance bias) Sick Leave	Unclear risk	Personal communication: "The allocation of the practice was known to patients eligible to participate. However, these patients did not know that there was another arm of the study that other practices participated in."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "We measured absenteeism at baseline, 6, 12, 18, and 24 months by calculating lost work hours from employee reports of how many full workdays and part workdays they missed due to illness or doctor visits, reflecting that employee reports demonstrate high agreement with employer records of absenteeism."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression was measured by self-report (CESD-D) and patients were not blinded to treatment allocation
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up at endpoint is considered to be high (27%). Risk of attrition bias was therefore deemed high although analyses accounted sufficiently for missing data according to authors: "Because analysis of missing data patterns produced no evidence of non ignorable missingness, we present unweighted models, noting that

Rost 2004 (*Continued*)

		weighted models produce closely comparable results."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up at endpoint is considered to be high (27%). Risk of attrition bias was therefore deemed high although although analyses accounted sufficiently for missing data according to authors: "Because analysis of missing data patterns produced no evidence of non ignorable missingness, we present unweighted models, noting that weighted models produce closely comparable results."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Schene 2006

Methods	RCT, two conditions; Regular referrals (including from occupational physicians), 48 weeks treatment. Follow up: 42 months. Lost to follow up at 12 months: T1: 13%; T2: 3%; at 42 months: T1: 25%; T2: 20%
Participants	62 were randomised (T1:32; T2:30). Setting: outpatient unit of Psychiatric department of Academic hospital. The Netherlands Inclusion:18 years; MDD (single episode or recurrent); BDI score >15; Work absenteeism due to depression of at least 50% of regular hours worked per week with a duration between 10 weeks and 2 years;Clinically estimated contribution of work to the onset and/or continuation of depression of > 50% of supposed causal factors Exclusion: MDD with psychotic features;history of psychosis, manic, hypomanic, or cyclothymic features; history of active drug or alcohol abuse or dependence; personality disorder according to DSM-IV Age: T1: 45.2 (SD 7.5); T2: 46.6 (SD 7.4) Female: T1: 53%; T2: 50% Married: T1: 63%; T2: 53% Mean hours employment: T1: 36.5 (SD 10.4); T2: 36.4 (SD 7.8)
Interventions	T1: Treatment as usual (TAU) following evidence-based guidelines (APA Guideline) ;This consisted of clinical management according to APA Guideline and antidepressants, if indicated and accepted by patients, according to our standardized stepwise drug treatment regimen or algorithm. Visits consisted of symptom assessment, psycho-education, general support and cognitive behavioral techniques, and if indicated medication prescription, dose titration and review of adverse effects. In case of any clinical significant deterioration in condition patients could be referred for partial or full-time hospitalisation within the Program. Patients were treated by three supervised senior psychiatric residents. Visits regularly took 30 minutes every 2 to 4 weeks T2: Treatment as usual + occupational therapy (TAU + OT) TAU plus occupational

	therapy (OT): same outpatient treatment; OT: diagnostic phase (4 weeks): occupational history, video observation in a role-played work situation, contact with occupational physician of patient's employer and written conclusions including a plan for work reintegration therapeutic phase (24 weeks): this phase had three sub-phases: preparation of work reintegration, contacting the working place and if possible starting to work. In the individual sessions these three phases were followed: further analyses of the relationship between work and depression, exploration of work problems, and support and evaluation of work resume. Specific individual issues from the group sessions were elaborated. The first half of these two-hour group sessions were spend on discussing and exchanging individual progress. In the second half seven themes were successively discussed: being passive, stress on the work place, personal bounds and limits, powerful and powerless, perfectionism, conflicts and prevention. Patients were treated by three supervised senior psychiatric residents. + two occupational therapists diagnostic phase (4 weeks): 5 visits therapeutic phase (24 weeks): 24 weekly group sessions (8-10 patients) and 12 individual sessions (45 minutes) follow-up phase (20 weeks): 3 individual visits
Outcomes	Absenteeism: 1) total number of hours worked during 6-month periods up to 42nd month (primary outcome) 2) proportion of patients working at least 1 hour per week 3) proportion of patients working at least 16 hours per week 4) time from T1 to partial or full return to work Clinical: 1) % meeting DSM IV criteria at 6/42 months 2) change in BDI at 6/42 months 1) depression according to DSM-IV at 12 months 2) change in BDI-score (baseline-12 months)
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Patients who met the inclusion criteria were randomly assigned to TAU or TAU +OT in blocks of 20 by use of computer-generated cards stored as concealed assignment codes in consecutively number sealed envelopes under the responsibility of an independent research associate."
Allocation concealment (selection bias)	Low risk	"Patients who met the inclusion criteria were them randomly assigned to TAU or TAU +OT in blocks of 20 by use of computer-generated cards stored as concealed

Schene 2006 (*Continued*)

		assignment codes in consecutively number sealed envelopes under the responsibility of an independent research associate."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Risk of performance bias considered high as the TAU cannot be considered equally desirable as TAU +OT for patients. Personal communication: "patients and clinical personnel were not blinded."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "Work resumption data were assessed by a study-specific questionnaire at T2, T3, T4 and T5."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	"Depression was assessed by the BDI, a self-report measure of severity of depressive symptoms." Patients were not blinded to treatment allocation
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up was high: T1: 25%; T2: 20%. Risk of attrition bias was therefore deemed high even though appropriate imputation methods have been used: "Complete T4 data were obtained on 28 (88%) of TAU patients and on 29 (97%) of TAU +OT patients. For T5 these figures were 24 (75%) for TAU and 24 (80%) for TAU + OT." "Both GEE and Proc Mixed give unbiased effect estimates taking into account missing data."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up was high: T1:25%; T2: 20%. Risk of attrition bias was therefore deemed high even though appropriate imputation methods have been used: "Complete T4 data were obtained on 28 (88%) of TAU patients and on 29 (97%) of TAU +OT patients. For T5 these figures were 24 (75%) for TAU and 24 (80%) for TAU +OT." "Both GEE and Proc Mixed give unbiased effect estimates taking into account missing data."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk

Schene 2006 (*Continued*)

Other bias	Unclear risk	None identified

Schoenbaum 2001

Methods	RCT with randomisation on the level of clinic. Clinic clusters were matched based on patient demographics, clinician specialty, and distance to mental health providers. Recruitment: study staff screened consecutive patient visitors. Follow up: 24 months. Lost to follow-up: T1: 15%; T2: 13%
Participants	1356 were randomised (T1:913; T2: 443). Setting: 46 Primary care clinics in 6 community-based managed care organisations in the US Inclusion: depressed, intend to use clinic for next 12 months; Probable depressive disorder:at least 2-weeks depressed mood or loss of interest in last year or persistent over year + at least 1 week depression in last 30 days Exclusion: < 18 years, acute medical emergency, did not speak English or Spanish, no insurance or public pay arrangement that covered care delivered by mental health specialists Age: T1: 44.5 (SD15.5); T2: 42.2 (SD 13.9) Female: T1: 74%; T2: 69% Married: T1: 54%; T2: 55%
Interventions	T1: Quality improvement program (QI meds or QI therapy). Treatment type or content Quality improvement (QI) program: practices were provided with training and resources to initiate and monitor QI programs according to local practice goals and resources. For both interventions (QI-meds and QI therapy): local practice teams were trained in a 2-day workshop to provide clinician education and to supervise intervention staff. Practice nurses were trained as depression specialists, following a written protocol, to assist in initial patient assessment, education and motivation for treatment. Practice teams were given patient education pamphlets and videotapes, patient tracking forms, and clinician manuals and pocket reminder cards and were encouraged to distribute them. The materials described guideline-concordant care and described antidepressant medication and psychotherapy as equally effective. In both conditions resources were made available to obtain specific form of therapy (medication or psychotherapy) For QI-meds: nurse specialists were trained to support medication adherence through monthly telephone contacts or visits for 6 or 12 months, randomised at patient level In QI-therapy: practice therapists were trained to provide individual and group CBT, following a protocol T2: Usual care: mailing of practice guidelines to primary care professionals
Outcomes	Absenteeism: 1) days worked during 24 months follow-up for whole sample 2) number of reported sick days for employed subsample in previous 4 weeks at each 6 months period Clinical: 1) CES-D
Notes	

Schoenbaum 2001 (*Continued*)

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Within blocks, we used a random number table to assign clusters to usual care or QI interventions."
Allocation concealment (selection bias)	High risk	Randomisation was on the level of practice and primary care clinicians were not blinded for allocation during enrolment of patients
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Patients and personnel were not blinded: "We asked all primary care clinicians to enrolls prior to their knowledge of intervention status." "Patients learned of their intervention status after enrolment." Personal communication: "Subjects in the interviews were not blinded, but may or may not have known their intervention status given the nature of interventions." Interventions were not equally desirable for patients, so risk of performance bias high
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "We also examined days missed from work due to illness, which patients reported for the 4 weeks preceding each follow-up study. "
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression was measured by self-report and patients were not blinded to treatment allocation 'We assessed depressive symptoms at baseline and follow-up using a 23-item version of the Center of Epidemiologic Studies Depression (CES-D) Scale, developed by Daniel Ford. This version drops 6 items and adds others to approximate DSM-IV criteria. Items responses were summed."
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Lost to follow up for the depressive symptoms is 15% but appropriate imputation methods have been used. "The data are weighted for the probability of study enrolment and follow-up response to the char-

Schoenbaum 2001 (*Continued*)

		acteristics of the eligible sample. We used multiple imputations for missing items at each wave."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Lost to follow for the economic survey is 15% but appropriate imputation methods have been used. "The data are weighted for the probability of study enrolment and follow-up response to the characteristics of the eligible sample. We used multiple imputations for missing items at each wave."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Simon 1998

Methods	RCT (consisting of 2 substudies) with two conditions. Recruitment: participating primary care physicians were asked to refer any adult outpatient initiating care for depression and willing to consider treatment with antidepressant medication. The research assistant screened for eligibility Follow up: 7 months. Lost to follow up: substudy 1: 15%; substudy 2: 23%
Participants	156 patients with MDD were randomised (T1: 80; T2: 76). Setting: Large primary care clinic in managed care setting in US Inclusion: diagnosis definite or probable major depression by primary care physician; Agreed to antidepressant medication; SCL-score of at least 0.75; Age 18 to 80 yrs Exclusion: current alcohol abuse (score at least 2 CAGE questionnaire); current psychotic symptoms or serious suicidal ideation or plan; dementia; pregnancy; terminal illness; limited; command of English; plan to disenrol from insurance plan within 12 months Age: substudy1: T1: 43.2 (SD 15.4); T2: 42.3 (SD 12.7); substudy2: T1: 43.1 (SD 9.3); T2: 44.8 (SD 15.9) Female: substudy1: T1: 78%; T2: 88%; substudy: 77%; T2: 74% Married or cohabiting: substudy1: T1: 47%; T2: 55%; substudy2: T1: 48%; T2: 32% Employed: substudy 1: T1: 71%; T2: 63% substudy 2: T1: 87%; T2: 74%
Interventions	T1: Multifaceted intervention. Goal: increase likelihood that treatment would be conform primary care depression guidelines Components: (1) written and videotaped patient education material (2) increased frequency of follow-up visits during first 8 weeks (3) advice to physicians regarding changes in pharmacotherapy (4) monitoring of medication side-effects, medication adherence, treatment response and follow-up visits frequency by study staff to treating physician substudy1, psychiatrist-liaison collaborative intervention: (a) co-management by consulting psychiatrist and physicians during first 6 weeks of treatment, (b) 1 week after start treatment all patients attended an extended structured visit with physician to review symptoms, barriers to adherence, side-effects, and goals for

	behavioural activation. (c) after 2 weeks: consultation with study psychiatrist discussing treatment response and medication (adjustment if needed), (d) week 3 physician visit, (e) week 4 psychiatrist visit (f) monthly case conferences between psychiatrist and physician substudy 2, psychologist-liaison collaborative intervention: Standardised brief psychotherapy program. Face-to-face psychiatric consultation on as-needed basis. Components psychotherapy: (a) education, skills training, and written homework (b) interventions to enhance medication adherence (c) behavioural activation and (d) brief cognitive interventions. Weekly meetings between therapists and study psychiatrists. Study clinicians communicated with physicians throughout study about progress and changes in medication psychotherapy: 4-6 visits over 6 weeks (total time 2,5 to 3,5 hour) Telephone contacts at 2, 4, 12 and 24 weeks after last face-to-face session T2: Usual primary care. Could include any service normally available including pharmacotherapy, referral to mental health service or self-referral to non-GHC services
Outcomes	Absenteeism: 1) % unable to work due to illness 2) n of days of missed work or school out of last 90 for employed subsample Clinical: 1) proportion of patients with MDD who experienced at least 50% reduction in depressive symptoms on IDS 2) SCL for employed subsample 3) IDS for employed subsample
Notes	Data are provided for subgroup of MDD only, both substudies combined

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Patients were randomly assigned using computer generated random numbers."
Allocation concealment (selection bias)	Low risk	Personal communication: "The primary care physicians or the research assistant did not know anything about the randomization status of the next patient. Randomization was performed 1-7 days after the baseline assessment by the study manager."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Personal communication: "Patient participants and their treating clinicians were not blinded - and it would not have been possible to do so." Interventions were not equally desirable for patients, so risk of performance bias high

Simon 1998 (*Continued*)

Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "One of the four assessments included questions adapted from the National Health Interview Survey regarding days of missed work or school due to health."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	"Folllow-up telephone interviewers were blinded to treatments assignment." "Two of the assessments included a 20-item depression scale extracted from the Hopkins Symptom Checklist or SCL and a version of the clinician-rated Inventory of Depressive Symptoms or IDS modified for telephone administration."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Lost to follow-up is considered to be high: T1: 17%; T2: 21%. Risk of attrition bias was therefore deemed high although appropriate imputation methods have been used: "Model were estimated using generalized estimating equations (GEE) to account for multiple assessments and to allow for missing data"
Incomplete outcome data (attrition bias) Sick Leave	High risk	Lost to follow-up is considered to be high: T1: 17%; T2: 21%. Risk of attrition bias was therefore deemed high although appropriate imputation methods have been used: "Model were estimated using generalized estimating equations (GEE) to account for multiple assessments and to allow for missing data"
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Vlasveld 2013

Methods	RCT. Recruitment: 22 months. Follow up: 12 months. Lost to follow up: 41.3%
Participants	126 were randomised (T1:65; T2:61); Setting: the study was carried out within a large occupational health service in the Netherlands Inclusion: workers on sickness absence between 4 and 12 weeks, whose absence was diagnosed by occupational physicians (OPs) as due to mental disorder, who screened

	positively for depressive disorder (i.e. score ≥ 10 on 9-item 0 to 27 depression subscale of Patient Health Questionnaire), who have informed consent and who met the DSM-IV criteria for MDD and gave written informed consent Exclusion: workers who were suicidal, psychotic or had a primary diagnosis of substance abuse or dependence, as assessed by the MINI Age: T1: 43.4 (SD 11.4); T2: 41.9 (SD 11.4) Male: T1: 45.9%; T2: 46.2% Marital status: T1: 73.3% married or cohabiting; T2: 60.0% married or cohabiting Educational level: T1: 35.0% high; T2: 36.1% high; T1 30.0% average; T2: 36.0% average; T1: 35.0% low; T2: 27.9% low Dutch nationality: T1: 91.8%; T2: 95.4%
Interventions	T1: Collaborative care intervention. Provided by the Occupational Physician Care Manager (OP-CM), contained the following elements: 6 to 12 sessions of Problem Solving Therapy, manual-guided self-help, a workplace intervention and, depending on patient preference, prescription of antidepressant medication according to a treatment algorithm. In order to enhance the adherence to the treatment model, ongoing supervision and psychiatric consultation was provided to the OP-CMs. Also, a web-based tracking system was developed to support the OP-CM in monitoring treatment outcomes and in adhering to the stepped care protocol. In case of questions regarding the treatment, prescription of antidepressants, or (lack of) progress of the worker, the OP-CM was prompted by the web-based tracking system to consult the psychiatrist T2: Usual care. Sick-listed workers start to visit the company's OP before the 6th week of sickness absence. The guidance of company's OP is protocolised according to the OP guidelines of the Dutch Board for Occupational Medicine. In practice, whether or not sick-listed workers will receive treatment for MDD may vary considerable. The actual care that was provided was assessed by questionnaires in both groups
Outcomes	Absenteeism: 1) the duration until lasting, full RTW. The duration until lasting, full RTW was defined as the duration of sickness absence due to MDD in calendar days, from the day of randomisation until full RTW for at least 4 weeks without partial or full recurrence 2) the total number of sickness absence days, calculated for the entire follow up Clinical: 1) severity of depression, assessed by the PHQ-9 2) time to first response on depressive symptoms. Response is defined as a reduction in depressive symptoms of at least 50% 3) time to first remission, defined as a score of less than 5 on the PHQ-9
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"The randomization scheme was prepared by a computer, with blocks of four, by an independent statistician."

Vlasveld 2013 (*Continued*)

Allocation concealment (selection bias)	Low risk	"While assessing eligibility for the study, both the research assistant and the participant were blinded for the allocation."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Participants were not blinded and both interventions were not equally desirable for them, so risk of performance bias was high. "Then, the participant was informed about the computer generated allocation status by the research assistant. Next, the baseline questionnaire was sent by mail."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Low risk as sickness absence data were based on registration database. "Sickness absence data were derived from the register of the occupational service 1 year after randomization."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Data about depressive symptoms were collected by a self-report questionnaire and patients were not blinded to treatment allocation
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Lost to follow up was high. "Lost to follow-up rates at 3,6, 9 and 12 months were respectively 22.2%, 28.6%, 33.3% and 41. 3%." Risk of attrition bias was considered high even though an appropriate method has been described to account for this missing data: 'If there is missing data on costs and/or effects, and the additional uncertainty it introduces, multiple imputation will be used."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	No missing sickness absence data
Selective reporting (reporting bias)	Low risk	The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way
Other bias	Unclear risk	None identified

Wade 2008

Methods	A double-blind, multinational randomised study. Recruitment: outpatients with MDD were recruited in psychiatric and general practice settings, from September 2005 to September 2006. Follow up: 24 weeks. Lost to follow up: 23% (clinical outcome) and 24.4% (sick leave)
Participants	295 were randomised (T1: 144; T2: 151). Setting: outpatients of 35 centra of psychiatric and general practice settings. Inclusion: patients with MDD (current episode assessed with the MINI), according to the DSM IV-TR criteria, outpatient of either sex, aged 18-65 years, with a MADRS total score ≥ 26 and a CGI-S score ≥ 4 at baseline visit. Patients with a secondary current comorbid anxiety disorder (DSM-IV TR criteria) could be included in the study, expect for obsessive-compulsive disorder, post-traumatic stress disorder, or panic disorder. Exclusion: if they met one or more of the DSM IV-TR criteria for any of the following: bipolar disorder, psychotic disorder or features, current eating disorders (anorexia nervosa, bulimia), mental retardation, any pervasive developmental disorder or cognitive disorder, or alcohol or drug abuse-related disorders within 12 months prior to baseline. In addition, patients at serious suicide risk, based on the investigator's clinical judgement, or who had a score of ≥ 5 on item 10 of the MADRS scale, were also excluded, as were those receiving formal behavioural therapy, or systematic psychotherapy, or were pregnant or breast feeding, or had a history of lactose intolerance. Patients with a history of hypersensitivity or non-response to citalopram, or escitalopram, or duloxetine, or with increased intra-ocular pressure, or at risk of acute narrow-angle glaucoma, were also excluded. Patients were also excluded if they were taking the following psychotropic drugs within 2 weeks prior to baseline or during the study: MAOI or RIMA, SSRI (fluoxetine within 5 weeks), SNRIs, and tricyclic antidepressants, tryptophan, psychoactive herbal remedies, any drug used for augmentation of antidepressant action or any other antidepressant drugs, oral antipsychotic and antimanic drugs (including lithium), or ECT (within 6 months), dopamine antagonists, any anxiolytics (including benzodiazepines), any anticonvulsant drug, serotonergic agonists, narcotic analgesics, cardiac glycosides, type 1c anti-arrhythmics, oral anticoagulants, cimetidine, potent inhibitors of CYP2C19, CYP1A2, or medicinal products with a narrow therapeutic index predominantly metabolised by CYP2D6 Female: T1: 73.8%; T2: 71.2% Age: T1: 43.3 (SD 11.6); T2: 44.5 (SD 11.0) Marital status: T1: 27.0% single; T2: 20.5% single T1: 50.4% married or living as a couple; T2: 50.7% married or living as a couple T1: 17.7% divorced or separated; T2: 25.3% divorced or separated T1: 5.0% widowed; T2: 3.4% widowed Level of education: T1: 5.0% no degree or diploma; T2: 4.1% no degree or diploma T1: 29.1% elementary school; T2: 26.0% elementary school T1: 43.3% high school; T2: 45.2% high school T1: 11.3% non-university degree; T2: 15.1% non university degree T1: 11.3% university; T2: 9.6% university Employment status: T1: 58.9% paid employment or self-employed; T2: 60.3% paid employment or self-employed

	T1: 15.6% unemployed; T2: 18.5% unemployed T1: 5.0% student; T2: 4.8% student T1: 6.4% non-working spouse; T2: 3.4% non-working spouse T1: 7.8% retired; T2: 10.3% retired T1: 6.4% other; T2: 2.7% other Occupational status: T1: 34.8% no data available; T2: 36.3% no data available T1: 6.5% manager or administrator; T2: 12.9% manager or administrator T1: 16.3% professional; T2: 15.1% professional T1: 10.9% associate professional; T2: 10.8% associate professional T1: 8.7% clerical worker/secretary; T2: 10.8% clerical worker/secretary T1: 26.1% skilled labourer or factory worker; T2: 17.2% skilled labourer or factory worker T1: 27.2% services/sales (retail); T2: 24.7% services/sales T1: 4.3% other; T2: 8.6% other
Interventions	T1: escitalopram (SSRI), 10 mg/day for the first 2 weeks, and 20 mg/day for the rest of the period T2: duloxetine (SNRI), 60 mg/day for the 24 weeks, in accordance with the recommendations in the package insert for duloxetine in the participating countries
Outcomes	Absenteeism: 1) percentage of patients taking sick leave 2) mean per patient sick leave duration in days Clinical: 1) adjusted mean change in the MADRS total score 2) MADRS total score 3) HAMD-17 4) remission, defined as MADRS $\leq$ 12 or post hoc as HAMD-17 $\leq$ 7) 5) response, defined as $\geq$50% decrease from baseline in MADRS or (post hoc) HAMD total score Work functioning: 1) impairment, assessed by the Sheehan Disability Scale
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Patients who met the selection criteria at the baseline visit were assigned to 24 weeks of double-blind treatment in a 1:1 ratio of escitalopram or duloxetine treatment according to a computer-generation randomization list." "At each study centre, sequentially enrolled patients were assigned to the lowest randomization number available in blocks of 4."

Allocation concealment (selection bias)	Low risk	"The details of the randomization series were unknown to any of the investigators and were contained in a set of sealed opaque envelopes."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	"All study personnel and participants were blinded to treatment assignment for the duration of the entire study."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Sick leave was assessed by physicians, who are blinded for allocation status
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	The MADRS and HAMD-17 are assessed by a doctor, who were blinded for allocation status
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Lost to follow up is considered to be high (23%). Risk of attrition bias was therefore deemed high and no appropriate method has been used to account for this missing data: "The primary endpoint was the adjusted mean change in MADRS total score from baseline to week 24, based on the intention-to-treat set (ITT), comprising all patients who took at least one valid post-baseline MADRS assessment, and using last-observation-carried-forward (LOCF) analysis."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Lost to follow up is considered to be high (24.4%) Risk of attrition bias was therefore deemed high and no appropriate method has been used to account for this missing data: "In cases of premature study withdrawal, patients were assigned zero sick leave for missing assessments."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

Wang 2007

Methods	RCT. Recruitment: occurred between January 2004 and February 2005 using a 2-phase procedure. Follow up: 12 months. Lost to follow up: 12.3%
Participants	604 were randomised (T1:304; T2:300); Setting: Participants included 604 depressed workers 18 years and older and enrolled in United Behavioural Health (UBH), a large managed behavioral health care company; Inclusion: Respondents with at least moderate depression (phase 1: K-6 $\geq$ 9; Phase 2: QIDS-SR $\geq$ 8); Exclusion: employees with lifetime bipolar disorder, substance disorder, recent mental health specialty care or suicidally Age: T1: 40.7 (SD 10.5); T2: 42.4 (SD 10.8) Female: T1: 70.7 %; T2: 77.0%% College graduates: T1: 38.0%; T2: 43.8% (24.6%)
Interventions	T1: The structured telephone intervention: telephone outreach and care management program. Systematically assessed needs for treatment, facilitated entry into in-person treatment (both psychotherapy and antidepressant medication), monitored and supported treatment adherence, and (for those declining in-person treatment) provided a structured psychotherapy intervention by telephone. Intervention participants declining in-person treatment and experiencing significant depressive symptoms after 2 months were offered a structured 8-session cognitive behavioural psychotherapy program T2: Usual care. Patients were advised to consult a clinician and could receive any normally available insurance benefit or service (eg, psychotherapy or pharmacotherapy), just not the additional telephone care management components provided to those in the intervention group
Outcomes	Absenteeism: 1) actual weekly hours worked among the employed, assessed by Health and Productivity Questionnaire (HPQ), a validated self-report instrument Clinical: 1) depression severity, assessed by QIDS-SR Functioning: 1) on-the-job performance, assessed by HPQ
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Randomization was carried out by the survey research firm conducting eligibility assessments with a computerized procedure that classified respondents for eligibility and used a random number generator to assign participants to intervention or usual care."
Allocation concealment (selection bias)	Low risk	"Patient treatment allocation was concealed."

Blinding of participants and personnel (performance bias) Sick Leave	High risk	Participants were not blinded. "Participants were advised not to offer information information to their interviewers regarding their intervention status." Interventions not equally desirable for both groups which entails a high risk of bias. "Respondents were told they might be invited to participate in an innovative treatment program."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	HPQ is a self-report instrument. As patients were aware of their allocation status, risk of bias high
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	QID-SR is a self-report instrument. As patients were aware of their allocation status, risk of bias high
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Lost to follow up: T1: 14.5%; T2: 10% but appropriate method has been used to account for missing data: "Multiple imputation was used to adjust for some participants not completing either 6-months (35 intervention and 22 usual care) or 12 month (44 intervention and 30 usual care) interviews." "Intervention effects on depression severity were estimated using multiple imputation linear regression with simulated standard errors."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Lost to follow up: T1: 14.5%; T2: 10% but appropriate method has been used to account for missing data: "Multiple imputation was used to adjust for some participants not completing either 6-months (35 intervention and 22 usual care) or 12 month (44 intervention and 30 usual care) interviews." "Comparable multiple imputation regression analyses were used to estimate intervention effects on work outcomes."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Unclear risk	None identified

BDI = Back Depression Inventory

CAGE = The name of which is an acronym of its four questions, is a widely used method of screening for alcoholism
CAU = Care as usual
CES-D = Center for Epidemiologic Studies Depression scale
CMD = Common mental disorders
CMHN = Community Mental Health Nursing
CIS-R = Clinical Interview Schedule-Revised
CTU = Copenhagen Trial Unit
DSM-IV = Diagnostic and Statistical Manual of Mental Disorders
4DSQ = Four-Dimensional Symptom Questionnaire
EAP = Employee Assistance Programme
ECT = Electroconvulsive therapy
FDA = Food and Drug Administration
GAS = Global Assessment Scale
GCI = Clinical Global Impression Scale
GEE = Generalized Estimating Equation
GP = General practitioner
GHQ-12 = General Health Questionnaire
HADS(-D)= Hospital Anxiety en Depression Scale
HAMD-D(17) = Hamilton Rating Scale for Depression
HDRS = Hamilton Rating Scale for Depression
HPQ = Health and Work Performance Questionnaire
HRSD = Hamilton Rating Scale for Depression
ICD-10 = International Statistical Classification of Diseases and Related Health Problems
IDS = Inventory of Depressive Symptomatology
LOCF = Last Observation Carrierd Forward
MADRS = Montgomery-Asberg Depression Scale
MAO = Monoamine oxidase
MAOI = Monoamine oxidase inhibitor
MINI = Mini International Neuropsychiatric Interview
MOS-SF 36 = Medical Outcomes Study 36-Item Short Form Health Survey
MDD = Major depressive disorder
OP = Occupational Physician
OT = Occupational therapy
PHQ = Patient Health Questionnaire
PST = Problem Solving Therapy
QI = Quality improvement
QIDS-SR = Quick Inventory of Depressive Symptomatology-Self Report
RCT = Randomized controlled trial
RIMA = Reversible inhibitors of monoamine oxidase A
RTW = Return to work
RTW-E = Exposure based return to work program
SAS = Social Adjustment Scale
SCL = Symptom Checklist Score
SNRI = Selective Serotonin and Noradrenalin Reuptake Inhibitor
SSRI = Delective serotonin reuptake inhibitor
TAU = Treatment as usual
TCA = Tricyclic antidepressant
WLQ = Work Limitations Questionnaire
WHI = Work and Health Initiative

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Aelfers 2013	Participants are people with a mild to moderate depression
Ahola 2012	Sickness absence was not measured as outcome measure
Alexopoulos	No worker population and sickness absence not measured as outcome measure
Amore 2001	Sickness absence was not measured as outcome measure
Bakker 2007	Patients suffered from mental health problems, less than 50% of these are patients with a depressive disorder
Barbui 2009	Sickness absence was not measured as outcome measure
Bech 2000	It is a meta-analysis instead of a RCT
Becker 1998	Participants were people with severe mental ilness such as schizophrenia
Blonk 2007	Patients suffered from psychological complaints, including adjustment disorders. Patients with a major depression were excluded from the study
Boyer 1998	Sickness absence was not measured as outcome measure
Brandes 2011	Sickness absence was not measured as outcome measure
Brouwers 2007	It is meta-analysis instead of a RCT
Carlin 2010	Sickness absence was not measured as outcome measure
Castillo-Pérez 2010	Sickness absence was not measured as outcome measure
Dick 1985	This study took place in an inpatient care setting
Dunlop 2011	Sickness absence was not measured as outcome measure
Eklund 2012	No RCT but a matched-control design was used
Erkkilä 2011	Sickness absence was not measured as outcome measure
Finley 2003	Sickness absence was not measured as outcome measure
Folke 2012	This study is done in a sample of unemployed individuals
Forman 2012	Participants were students
Furukawa 2012	Participants with mild depression were included in this study; people with a major depressive disorder were excluded

(*Continued*)

Gournay 1995	Participants suffered from a range of non-psychotic symptoms, data for the depressed subgroup only could not be provided
Hackett 1987	Inclusion criterion in this study was 'clinical diagnosis of chronic muscle contraction headache'
Hirani 2010	Sickness absence was not measured as outcome measure
Hordern 1964	This study took place in a hospital setting
Knekt 2011	It is quasi-experimental study
Kojima 2010	Sickness absence was not measured as outcome measure
Kroenke 2001	Sickness absence was not measured as outcome measure
Kuhs 1996	Sickness absence was not measured as outcome measure
Lagerveld 2012	Major depressive disorder was excluded in this study
Lam 2012	Sickness absence was not measured as outcome measure
Lexis 2011	The focus in this study is on relatively mild complaints
Martinez 2011	Sickness absence was not measured as outcome measure
Meyer 2009	Sickness absence was not measured as outcome measure
Mino 2006	Prevention study; subjects were not depressed
Morgan 2011	Participants are people with sub-threshold depression
Mundt 2001	Sickness absence was not measured as outcome measure
Oakes 2012	Sickness absence was not measured as outcome measure
Salminen 2008	Sickness absence was not measured as outcome measure
Sandahl 2011	Sickness absence was not measured as outcome measure
Schmitt 2008	It is not a RCT but a review
Schoenbaum 2002	This study turned out to be a publication on the same study as Schoenbaum 2001 (which was also included)
Simon 2000	Sickness absence was not measured as outcome measure
Sir 2005	Sickness absence was not measured as outcome measure
Stant 2009	Sickness absence was not measured as outcome measure

(*Continued*)

Wells 2000	This trial is the basis of the economic evaluation of Schoenbaum 2001
Zambori 2002	Design was CCT instead of RCT
Zeeuw 2010	This study focuses on employees with minimal symptoms of depression

Characteristics of ongoing studies *[ordered by study ID]*

Beurden 2013

Trial name or title	Not yet assessed
Methods	Cluster RCT
Participants	common mental disorders
Interventions	guideline-based care by occupational physicians
Outcomes	Return-to-work
Starting date	2011
Contact information	Department of Social and Behavioral Sciences, Tranzo Scientific Center for Care and Welfare, Tilburg University, PO Box 90153, Tilburg, 5000 LE, the Netherlands
Notes	

Geraedts 2013

Trial name or title	Happy@Work
Methods	RCT
Participants	Employees with depressive symptoms
Interventions	Web-based guided self-help
Outcomes	Health and Work Performance Questionnaire
Starting date	Not yet assessed
Contact information	a.s.geraedts@vu.nl
Notes	

Heer 2013

Trial name or title	TCC: PAINDIP
Methods	Randomised placebo-controlled multicentre trial
Participants	Major depressive disorder (MDD) and (sub)chronic pain
Interventions	Transmural collaborative care with consultation letter (TCCCL) and duloxetine in collaboration with primary care
Outcomes	Not yet assessed
Starting date	Not yet assessed
Contact information	Not yet assessed
Notes	

Hellstrom 2013

Trial name or title	The effect of IPS-modified, an early intervention for people with mood and anxiety disorders: study protocol for a randomised clinical superiority trial. Trials 2013, 14:442. doi:10.1186/1745-6215-14-442
Methods	Randomized clinical superiority trial
Participants	Participants with mood and anxiety disorders recently employed or enrolled in education
Interventions	IPS-modified; an individualised supported employment intervention, aiming at supporting people with recently diagnosed anxiety or affective disorders to obtain and sustain competitive employment through mentor support
Outcomes	Competitive employment or education at 24 months
Starting date	Not yet assessed
Contact information	lone.hellstroem@regionh.dk
Notes	

Warmerdam 2007

Trial name or title	Not yet assessed
Methods	RCT with three conditions: two treatment conditions and one waiting list control group
Participants	Subjects with symptoms of depression (≥ 16 on the Center for Epidemiological Studies Depression scale) from the general population

Warmerdam 2007 (*Continued*)

Interventions	Two Internet-based treatments for depression, namely cognitive behavioural therapy and problem-solving therapy
Outcomes	Absence at work
Starting date	Not yet assessed
Contact information	eh.warmerdam@psy.vu.nl
Notes	Results at 12 weeks are published: https://ce1ul13jdba-qne8l9ebidp.sec.amc.nl/pmc/articles/PMC2629364/ This did not include the work outcome

DATA AND ANALYSES

Comparison 1. Work-directed plus clinical versus clinical alone (medium term)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	3	251	Std. Mean Difference (IV, Random, 95% CI)	-0.40 [-0.66, -0.14]
1.1 Occupational therapy plus CAU vs. CAU	2	179	Std. Mean Difference (IV, Random, 95% CI)	-0.30 [-0.61, 0.01]
1.2 Multi-component work-focused program vs. CAU	1	72	Std. Mean Difference (IV, Random, 95% CI)	-0.66 [-1.15, -0.16]
2 Depressive symptoms	3	251	Std. Mean Difference (IV, Random, 95% CI)	-0.32 [-0.88, 0.25]
2.1 Occupational therapy plus CAU vs. CAU	2	179	Std. Mean Difference (IV, Random, 95% CI)	-0.08 [-0.66, 0.50]
2.2 Multi-component work-focused program vs. CAU	1	72	Std. Mean Difference (IV, Random, 95% CI)	-0.84 [-1.34, -0.33]
3 Work functioning	2	189	Std. Mean Difference (IV, Random, 95% CI)	-0.31 [-0.79, 0.16]
3.1 Occupational therapy plus CAU vs. CAU	1	117	Std. Mean Difference (IV, Random, 95% CI)	-0.09 [-0.48, 0.29]
3.2 Multi-component work-focused program vs. CAU	1	72	Std. Mean Difference (IV, Random, 95% CI)	-0.58 [-1.08, -0.09]

Comparison 2. Work-directed plus clinical versus clinical alone (long term)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	2	179	Std. Mean Difference (IV, Random, 95% CI)	-0.19 [-0.49, 0.12]
1.1 Occupational therapy plus CAU vs. CAU	2	179	Std. Mean Difference (IV, Random, 95% CI)	-0.19 [-0.49, 0.12]
2 Depressive symptoms	1	117	Std. Mean Difference (IV, Random, 95% CI)	-0.63 [-1.02, -0.24]
2.1 Occupational therapy plus CAU vs. CAU	1	117	Std. Mean Difference (IV, Random, 95% CI)	-0.63 [-1.02, -0.24]
3 Work functioning	1	117	Std. Mean Difference (IV, Random, 95% CI)	-0.25 [-0.63, 0.14]
3.1 Occupational therapy plus CAU vs. CAU	1	117	Std. Mean Difference (IV, Random, 95% CI)	-0.25 [-0.63, 0.14]

Comparison 3. Work-directed plus clinical versus work-directed (medium term)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	1	126	Std. Mean Difference (IV, Random, 95% CI)	-0.14 [-0.49, 0.21]
1.1 Collaborative care vs. CAU	1	126	Std. Mean Difference (IV, Random, 95% CI)	-0.14 [-0.49, 0.21]
2 Depressive symptoms	1	74	Std. Mean Difference (IV, Random, 95% CI)	0.26 [-0.20, 0.72]
2.1 Collaborative care vs. CAU	1	74	Std. Mean Difference (IV, Random, 95% CI)	0.26 [-0.20, 0.72]

Comparison 4. Any work-directed versus alternative work-directed

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	1	75	Std. Mean Difference (IV, Random, 95% CI)	0.45 [-0.00, 0.91]
1.1 RTW-E vs. CAU	1	75	Std. Mean Difference (IV, Random, 95% CI)	0.45 [-0.00, 0.91]
2 Depressive symptoms	1	40	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.84, 0.49]
2.1 RTW-E vs. CAU	1	40	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-0.84, 0.49]

Comparison 5. Any antidepressant medication versus any other antidepressant medication

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	5		Std. Mean Difference (IV, Fixed, 95% CI)	Totals not selected
1.1 SSRI vs. SNRI	3		Std. Mean Difference (IV, Fixed, 95% CI)	0.0 [0.0, 0.0]
1.2 SSRI vs. TCA	1		Std. Mean Difference (IV, Fixed, 95% CI)	0.0 [0.0, 0.0]
1.3 SSRI vs. SSRI	1		Std. Mean Difference (IV, Fixed, 95% CI)	0.0 [0.0, 0.0]
2 Depressive symptoms	5	1514	Std. Mean Difference (IV, Random, 95% CI)	0.07 [-0.34, 0.48]
2.1 SSRI vs. SNRI	3	599	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.37, 0.73]
2.2 SSRI vs. TCA	1	635	Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2.3 SSRI vs. SSRI	1	280	Std. Mean Difference (IV, Random, 95% CI)	-0.23 [-0.47, 0.00]
3 Work functioning	1	635	Mean Difference (IV, Random, 95% CI)	-0.05 [-0.16, 0.06]
3.1 SSRI vs. TCA	1	635	Mean Difference (IV, Random, 95% CI)	-0.05 [-0.16, 0.06]

Comparison 6. Any antidepressant medication versus placebo

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	1	61	Std. Mean Difference (IV, Random, 95% CI)	0.48 [-0.05, 1.00]
1.1 TCA or MAO vs. placebo	1	61	Std. Mean Difference (IV, Random, 95% CI)	0.48 [-0.05, 1.00]
2 Work functioning	1	61	Std. Mean Difference (IV, Random, 95% CI)	-0.58 [-1.11, -0.05]
2.1 TCA or MAO vs. placebo	1	61	Std. Mean Difference (IV, Random, 95% CI)	-0.58 [-1.11, -0.05]

Comparison 7. Any psychological versus other psychological (medium term)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	1		Std. Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Short-term psychodynamic therapy vs. solution-focused therapy	1		Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
1.2 Long-term psychodynamic therapy vs. solution-focused therapy	1		Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Depressive symptoms	1		Std. Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1 Short-term psychodynamic therapy vs. solution-focused therapy	1		Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2.2 Long-term psychodynamic therapy vs. solution-focused therapy	1		Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3 Work functioning	1		Std. Mean Difference (IV, Random, 95% CI)	Totals not selected
3.1 Short-term psychodynamic therapy vs solution-focused therapy	1		Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3.2 Long-term psychodynamic therapy vs solution-focused therapy	1		Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]

Comparison 8. Any psychological versus other psychological (long term)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	1		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Short-term psychodynamic therapy vs. solution-focused therapy	1	36	Std. Mean Difference (IV, Random, 95% CI)	-0.91 [-1.62, -0.19]
1.2 Long-term psychodynamic therapy vs. solution-focused therapy	1	42	Std. Mean Difference (IV, Random, 95% CI)	-4.61 [-5.84, -3.39]
2 Depressive symptoms	1	263	Std. Mean Difference (IV, Random, 95% CI)	-1.85 [-2.99, -0.72]
2.1 Short-term psychodynamic therapy vs. solution-focused therapy	1	118	Std. Mean Difference (IV, Random, 95% CI)	-1.28 [-1.69, -0.86]
2.2 Long-term psychodynamic therapy vs. solution-focused therapy	1	145	Std. Mean Difference (IV, Random, 95% CI)	-2.44 [-2.90, -1.97]
3 Work functioning	1		Std. Mean Difference (IV, Random, 95% CI)	Totals not selected
3.1 Short-term psychodynamic therapy vs. solution-focused therapy	1		Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3.2 Long-term psychodynamic therapy vs. solution-focused therapy	1		Std. Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]

Comparison 9. Any psychological versus no intervention or care as usual

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	4		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Online/telephone CBT vs. CAU	3	326	Std. Mean Difference (IV, Random, 95% CI)	-0.23 [-0.45, -0.01]
1.2 CMHN vs. usual GP care	1	59	Std. Mean Difference (IV, Random, 95% CI)	0.22 [-0.36, 0.79]
2 Depressive symptoms	4		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Online/telephone CBT vs. CAU	3	408	Std. Mean Difference (IV, Random, 95% CI)	-0.56 [-0.76, -0.36]
2.2 CMHN vs. usual GP care	1	78	Std. Mean Difference (IV, Random, 95% CI)	0.22 [-0.31, 0.75]

Comparison 10. Psychological combined with antidepressant medication versus antidepressant medication alone

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	1	57	Std. Mean Difference (IV, Random, 95% CI)	-0.71 [-1.25, -0.17]
1.1 Psychodynamic therapy plus TCA vs. TCA	1	57	Std. Mean Difference (IV, Random, 95% CI)	-0.71 [-1.25, -0.17]
2 Work functioning or productivity	1	57	Std. Mean Difference (IV, Random, 95% CI)	-0.49 [-1.02, 0.04]
2.1 Psychodynamic therapy plus TCA vs. TCA	1	57	Std. Mean Difference (IV, Random, 95% CI)	-0.49 [-1.02, 0.04]
3 Depressive symptoms	1	74	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.57, 0.35]

Comparison 11. Psychological combined with antidepressant medication versus no intervention or usual care (medium term)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Enhanced primary care vs. CAU	2	969	Std. Mean Difference (IV, Random, 95% CI)	-0.02 [-0.15, 0.12]
1.2 Telephone outreach and care management program vs. CAU	1	604	Std. Mean Difference (IV, Random, 95% CI)	-0.21 [-0.37, -0.05]
2 Employment status	1	1356	Risk Ratio (M-H, Random, 95% CI)	1.08 [0.99, 1.18]
2.1 Enhanced primary care vs. CAU	1	1356	Risk Ratio (M-H, Random, 95% CI)	1.08 [0.99, 1.18]
3 Depressive symptoms	2	693	Std. Mean Difference (IV, Random, 95% CI)	-0.22 [-0.37, -0.07]
3.1 Enhanced primary care vs. CAU	1	89	Std. Mean Difference (IV, Random, 95% CI)	-0.14 [-0.56, 0.28]
3.2 Telephone outreach and care management program vs. CAU	1	604	Std. Mean Difference (IV, Random, 95% CI)	-0.23 [-0.39, -0.07]
4 Depressed yes/no	1	1356	Risk Ratio (M-H, Random, 95% CI)	0.89 [0.81, 0.98]
4.1 Enhanced primary care vs. CAU	1	1356	Risk Ratio (M-H, Random, 95% CI)	0.89 [0.81, 0.98]
5 Work functioning	1	604	Std. Mean Difference (IV, Random, 95% CI)	0.50 [0.34, 0.66]
5.1 Telephone outreach and care management program vs. CAU	1	604	Std. Mean Difference (IV, Random, 95% CI)	0.50 [0.34, 0.66]

Comparison 12. Exercise intervention versus no intervention or care as usual

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Days of sickness absence	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Supervised strength training vs. relaxation	1	65	Std. Mean Difference (IV, Random, 95% CI)	-1.11 [-1.68, -0.54]
1.2 Aerobic exercise vs. relaxation/stretching	2	180	Std. Mean Difference (IV, Random, 95% CI)	-0.06 [-0.36, 0.24]
2 Depressive symptoms	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Supervised strength training vs. relaxation	1	65	Std. Mean Difference (IV, Random, 95% CI)	0.15 [-0.39, 0.68]
2.2 Aerobic exercise vs. relaxation/stretching	2	180	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.12, 0.48]

Analysis 1.1. Comparison 1 Work-directed plus clinical versus clinical alone (medium term), Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 1 Work-directed plus clinical versus clinical alone (medium term)

Outcome: 1 Days of sickness absence

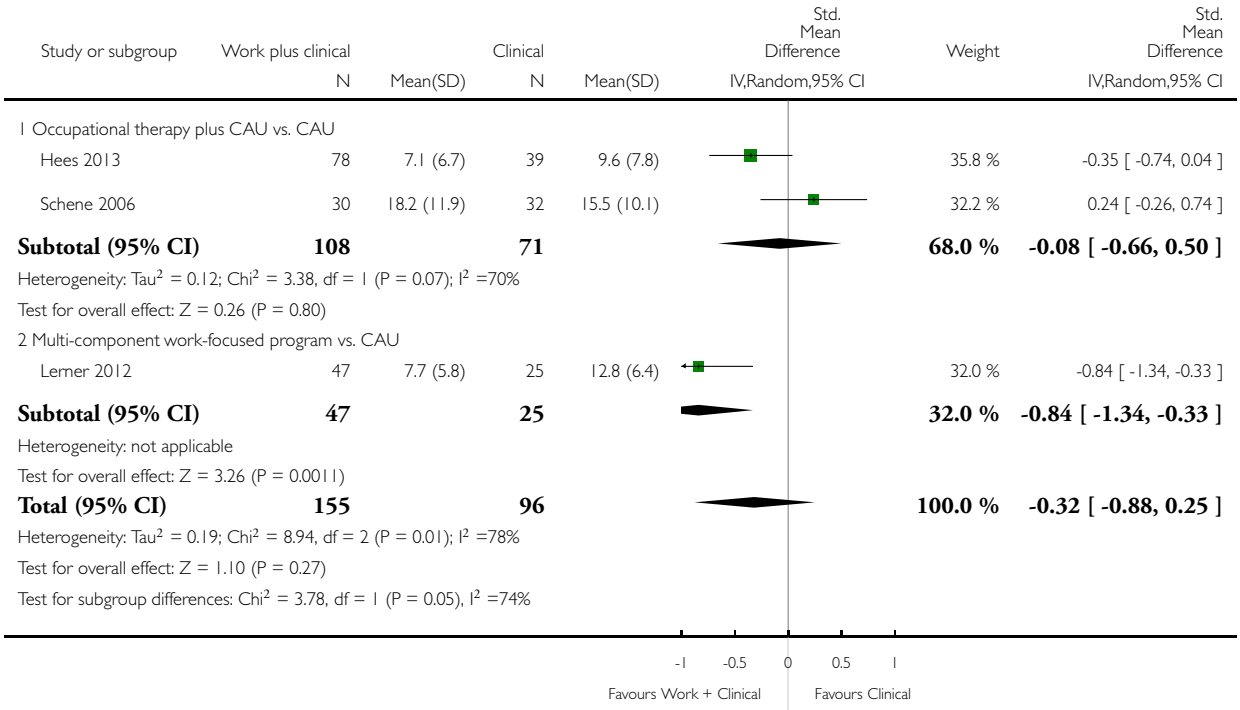
Study or subgroup	Work plus clinical N	Mean(SD)	Clinical N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Occupational therapy plus CAU vs. CAU							
Hees 2013	78	1.7625 (1.4875)	39	2.13 (1.6)		45.7 %	-0.24 [-0.62, 0.15]
Schene 2006	30	66.19 (40.36)	32	83.28 (41.79)		26.8 %	-0.41 [-0.91, 0.09]
Subtotal (95% CI)	**108**		**71**			**72.5 %**	**-0.30 [-0.61, 0.01]**
Heterogeneity: Tau² = 0.0; Chi² = 0.29, df = 1 (P = 0.59); I² =0.0%							
Test for overall effect: Z = 1.92 (P = 0.054)							
2 Multi-component work-focused program vs. CAU							
Lerner 2012	47	1 (1.2)	25	2.2 (2.6)		27.5 %	-0.66 [-1.15, -0.16]
Subtotal (95% CI)	**47**		**25**			**27.5 %**	**-0.66 [-1.15, -0.16]**
Heterogeneity: not applicable							
Test for overall effect: Z = 2.59 (P = 0.0096)							
Total (95% CI)	**155**		**96**			**100.0 %**	**-0.40 [-0.66, -0.14]**
Heterogeneity: Tau² = 0.0; Chi² = 1.72, df = 2 (P = 0.42); I² =0.0%							
Test for overall effect: Z = 2.99 (P = 0.0027)							
Test for subgroup differences: Chi² = 1.43, df = 1 (P = 0.23), I² =30%							

Analysis 1.2. Comparison 1 Work-directed plus clinical versus clinical alone (medium term), Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 1 Work-directed plus clinical versus clinical alone (medium term)

Outcome: 2 Depressive symptoms



Analysis 1.3. Comparison 1 Work-directed plus clinical versus clinical alone (medium term), Outcome 3 Work functioning.

Review: Interventions to improve return to work in depressed people

Comparison: 1 Work-directed plus clinical versus clinical alone (medium term)

Outcome: 3 Work functioning

Study or subgroup	Work plus clinical N	Mean(SD)	Clinical N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Occupational therapy plus CAU vs. CAU							
Hees 2013	78	33.1 (15)	39	34.7 (19.8)		55.3 %	-0.09 [-0.48, 0.29]
Subtotal (95% CI)	**78**		**39**			**55.3 %**	**-0.09 [-0.48, 0.29]**
Heterogeneity: not applicable							
Test for overall effect: Z = 0.48 (P = 0.63)							
2 Multi-component work-focused program vs. CAU							
Lerner 2012	47	26.5 (23.5)	25	40.1 (22.1)		44.7 %	-0.58 [-1.08, -0.09]
Subtotal (95% CI)	**47**		**25**			**44.7 %**	**-0.58 [-1.08, -0.09]**
Heterogeneity: not applicable							
Test for overall effect: Z = 2.31 (P = 0.021)							
Total (95% CI)	**125**		**64**			**100.0 %**	**-0.31 [-0.79, 0.16]**
Heterogeneity: Tau^2 = 0.07; Chi^2 = 2.34, df = 1 (P = 0.13); I^2 =57%							
Test for overall effect: Z = 1.29 (P = 0.20)							
Test for subgroup differences: Chi^2 = 2.34, df = 1 (P = 0.13), I^2 =57%							

Analysis 2.1. Comparison 2 Work-directed plus clinical versus clinical alone (long term), Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 2 Work-directed plus clinical versus clinical alone (long term)

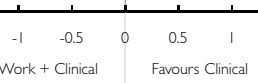
Outcome: 1 Days of sickness absence

Study or subgroup	Work plus clinical N	Mean(SD)	Clinical N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Occupational therapy plus CAU vs. CAU							
Hees 2013	78	1.3 (1.5625)	39	1.49 (1.5375)		62.9 %	-0.12 [-0.50, 0.26]
Schene 2006	30	49.81 (50.21)	32	65.5 (52.65)		37.1 %	-0.30 [-0.80, 0.20]
Total (95% CI)	**108**		**71**			**100.0 %**	**-0.19 [-0.49, 0.12]**

Heterogeneity: $Tau^2 = 0.0$; $Chi^2 = 0.32$, df = 1 (P = 0.57); $I^2 = 0.0\%$

Test for overall effect: Z = 1.20 (P = 0.23)

Test for subgroup differences: Not applicable

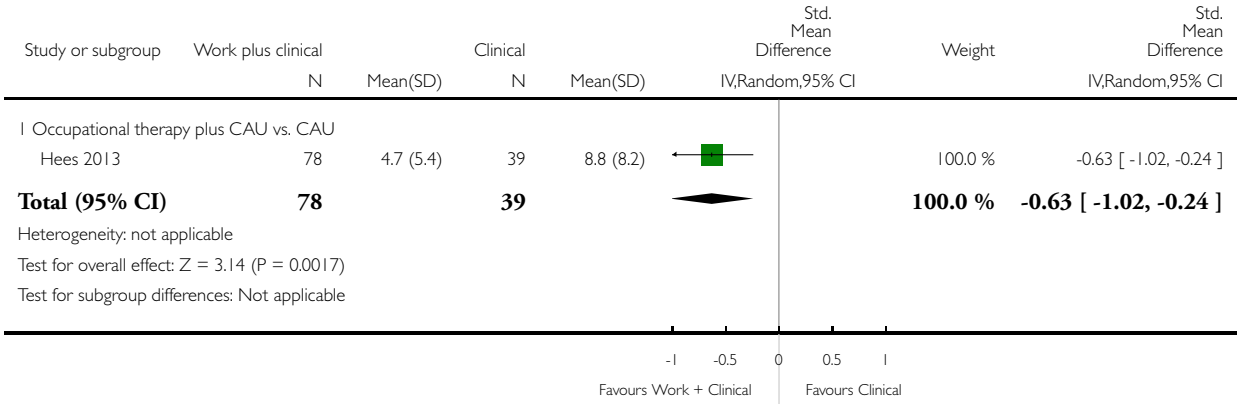


Analysis 2.2. Comparison 2 Work-directed plus clinical versus clinical alone (long term), Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 2 Work-directed plus clinical versus clinical alone (long term)

Outcome: 2 Depressive symptoms

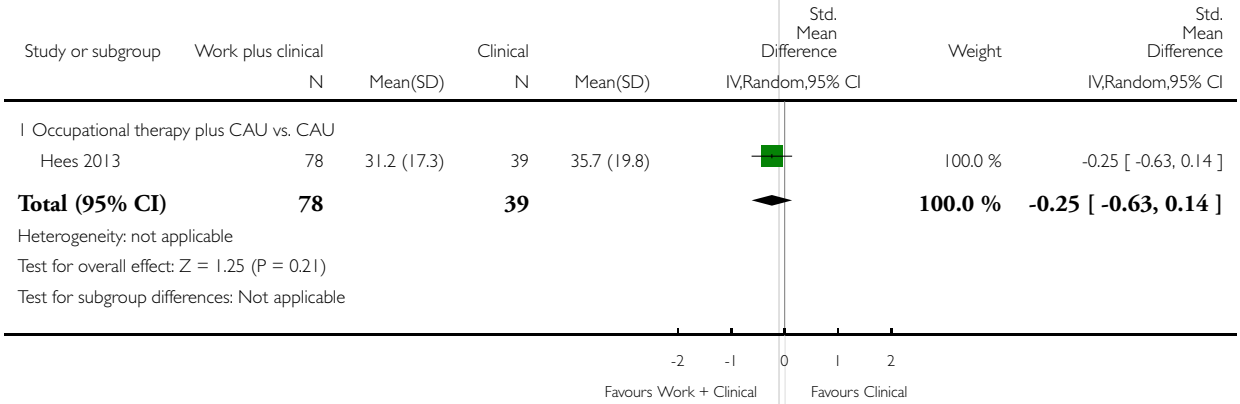


Analysis 2.3. Comparison 2 Work-directed plus clinical versus clinical alone (long term), Outcome 3 Work functioning.

Review: Interventions to improve return to work in depressed people

Comparison: 2 Work-directed plus clinical versus clinical alone (long term)

Outcome: 3 Work functioning



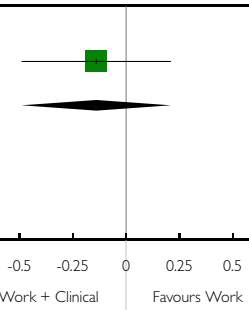
Analysis 3.1. Comparison 3 Work-directed plus clinical versus work-directed (medium term), Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 3 Work-directed plus clinical versus work-directed (medium term)

Outcome: 1 Days of sickness absence

Study or subgroup	Work plus clinical N	Mean(SD)	Work N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Collaborative care vs. CAU							
Vlasveld 2013	65	198 (120)	61	215 (118)		100.0 %	-0.14 [-0.49, 0.21]
Total (95% CI)	**65**		**61**			**100.0 %**	**-0.14 [-0.49, 0.21]**
Heterogeneity: not applicable							
Test for overall effect: Z = 0.80 (P = 0.43)							
Test for subgroup differences: Not applicable							

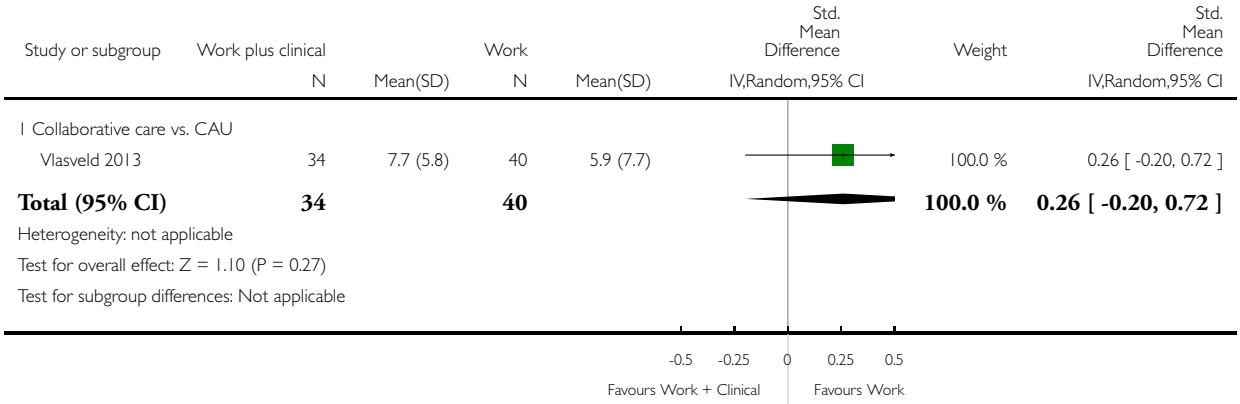


Analysis 3.2. Comparison 3 Work-directed plus clinical versus work-directed (medium term), Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 3 Work-directed plus clinical versus work-directed (medium term)

Outcome: 2 Depressive symptoms

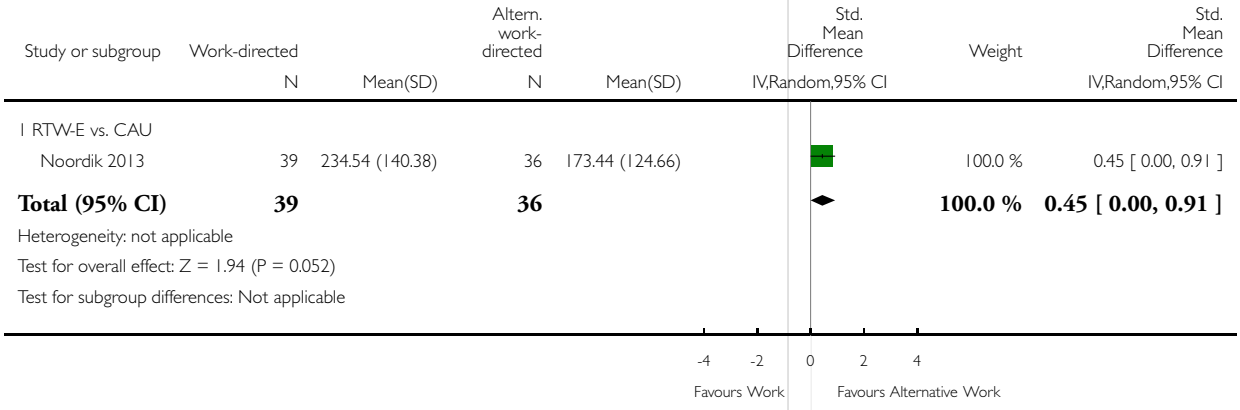


Analysis 4.1. Comparison 4 Any work-directed versus alternative work-directed, Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 4 Any work-directed versus alternative work-directed

Outcome: 1 Days of sickness absence



Analysis 4.2. Comparison 4 Any work-directed versus alternative work-directed, Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 4 Any work-directed versus alternative work-directed

Outcome: 2 Depressive symptoms

Study or subgroup	Work-directed N	Mean(SD)	Altern. work-directed N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 RTW-E vs. CAU							
Noordik 2013	13	0.46 (0.97)	27	0.85 (2.54)		100.0 %	-0.18 [-0.84, 0.49]
Total (95% CI)	**13**		**27**			**100.0 %**	**-0.18 [-0.84, 0.49]**

Heterogeneity: not applicable

Test for overall effect: Z = 0.52 (P = 0.60)

Test for subgroup differences: Not applicable

Analysis 5.1. Comparison 5 Any antidepressant medication versus any other antidepressant medication, Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 5 Any antidepressant medication versus any other antidepressant medication

Outcome: 1 Days of sickness absence

Study or subgroup	Antidepressant N	Mean(SD)	Other antide-pressant N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Std. Mean Difference IV,Fixed,95% CI
1 SSRI vs. SNRI						
Fernandez 2005	62	12.37 (21.51)	73	12.97 (22.52)		-0.03 [-0.37, 0.31]
Romeo 2004	49	28 (32)	45	19 (32)		0.28 [-0.13, 0.69]
Wade 2008	83	23.4 (27.87)	88	41.7 (35.41)		-0.57 [-0.88, -0.26]
2 SSRI vs. TCA						
Miller 1998	426	0.4 (1.98)	209	0.23 (2.26)		0.08 [-0.08, 0.25]
3 SSRI vs. SSRI						
Fantino 2007	138	11.56 (2)	142	12.18 (2)		-0.31 [-0.54, -0.07]

Analysis 5.2. Comparison 5 Any antidepressant medication versus any other antidepressant medication, Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 5 Any antidepressant medication versus any other antidepressant medication

Outcome: 2 Depressive symptoms

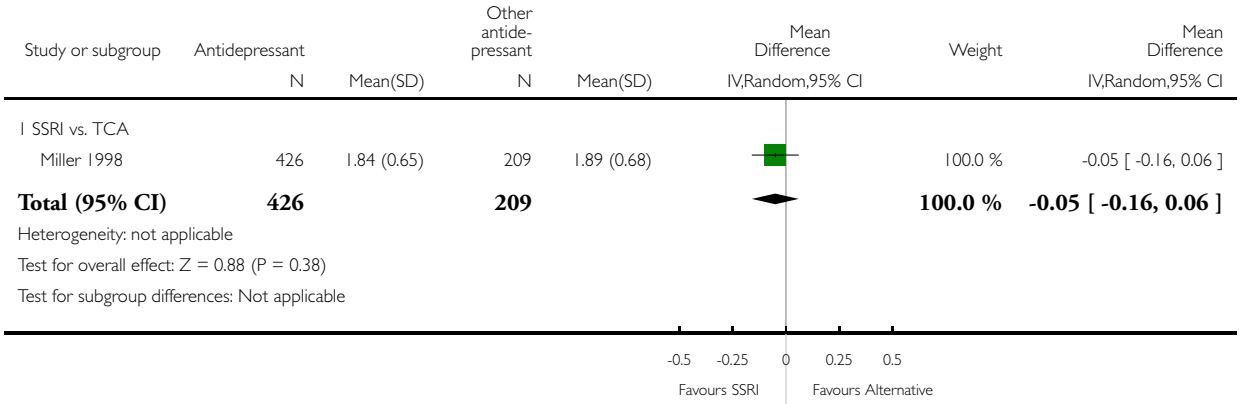
Study or subgroup	Antidepressant N	Mean(SD)	Other antide-pressant N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 SSRI vs. SNRI							
Fernandez 2005	62	5.95 (4.78)	73	5.79 (6.13)		23.7 %	0.03 [-0.31, 0.37]
Romeo 2004	84	16.42 (7.63)	93	10.99 (7.58)		24.5 %	0.71 [0.41, 1.02]
Wade 2008	141	8.89 (8.9)	146	10.58 (8.9)		25.9 %	-0.19 [-0.42, 0.04]
Subtotal (95% CI)	**287**		**312**			**74.1 %**	**0.18 [-0.37, 0.73]**
Heterogeneity: $Tau^2 = 0.21$; $Chi^2 = 21.60$, df = 2 (P = 0.00002); $I^2 = 91\%$							
Test for overall effect: Z = 0.64 (P = 0.52)							
2 SSRI vs. TCA							
Miller 1998	426	12.6 (0)	209	12.9 (0)			Not estimable
Subtotal (95% CI)	**426**		**209**				**Not estimable**
Heterogeneity: not applicable							
Test for overall effect: not applicable							
3 SSRI vs. SSRI							
Fantino 2007	138	13.9 (6.37)	142	15.4 (6.37)		25.9 %	-0.23 [-0.47, 0.00]
Subtotal (95% CI)	**138**		**142**			**25.9 %**	**-0.23 [-0.47, 0.00]**
Heterogeneity: not applicable							
Test for overall effect: Z = 1.96 (P = 0.050)							
Total (95% CI)	**851**		**663**			**100.0 %**	**0.07 [-0.34, 0.48]**
Heterogeneity: $Tau^2 = 0.16$; $Chi^2 = 27.38$, df = 3 (P<0.00001); $I^2 = 89\%$							
Test for overall effect: Z = 0.34 (P = 0.74)							
Test for subgroup differences: $Chi^2 = 1.85$, df = 1 (P = 0.17), $I^2 = 46\%$							

Analysis 5.3. Comparison 5 Any antidepressant medication versus any other antidepressant medication, Outcome 3 Work functioning.

Review: Interventions to improve return to work in depressed people

Comparison: 5 Any antidepressant medication versus any other antidepressant medication

Outcome: 3 Work functioning

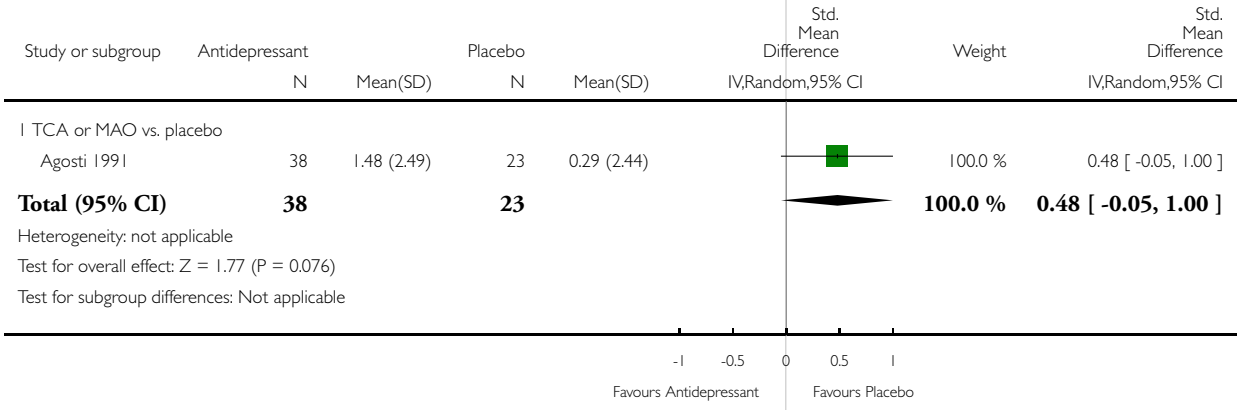


Analysis 6.1. Comparison 6 Any antidepressant medication versus placebo, Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 6 Any antidepressant medication versus placebo

Outcome: 1 Days of sickness absence



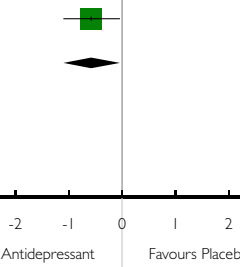
Analysis 6.2. Comparison 6 Any antidepressant medication versus placebo, Outcome 2 Work functioning.

Review: Interventions to improve return to work in depressed people

Comparison: 6 Any antidepressant medication versus placebo

Outcome: 2 Work functioning

Study or subgroup	Antidepressant N	Mean(SD)	Placebo N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 TCA or MAO vs. placebo							
Agosti 1991	38	1.8 (1.4)	23	2.6 (1.3)		100.0 %	-0.58 [-1.11, -0.05]
Total (95% CI)	**38**		**23**			**100.0 %**	**-0.58 [-1.11, -0.05]**
Heterogeneity: not applicable							
Test for overall effect: Z = 2.15 (P = 0.032)							
Test for subgroup differences: Not applicable							



Analysis 7.1. Comparison 7 Any psychological versus other psychological (medium term), Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 7 Any psychological versus other psychological (medium term)

Outcome: 1 Days of sickness absence

Study or subgroup	Psychodynamic N	Mean(SD)	Solution-focused N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Std. Mean Difference IV,Random,95% CI
1 Short-term psychodynamic therapy vs. solution-focused therapy						
Knekt 2013	34	2.79 (0.85)	13	2.58 (0.77)		0.25 [-0.39, 0.89]
2 Long-term psychodynamic therapy vs. solution-focused therapy						
Knekt 2013	38	4.18 (1.5)	13	2.58 (0.77)		1.16 [0.49, 1.83]

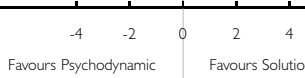
Analysis 7.2. Comparison 7 Any psychological versus other psychological (medium term), Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 7 Any psychological versus other psychological (medium term)

Outcome: 2 Depressive symptoms

Study or subgroup	Psychodynamic N	Mean(SD)	Solution-focused N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Std. Mean Difference IV,Random,95% CI
1 Short-term psychodynamic therapy vs. solution-focused therapy						
Knekt 2013	91	9.5 (1)	45	10.7 (1)		-1.19 [-1.58, -0.81]
2 Long-term psychodynamic therapy vs. solution-focused therapy						
Knekt 2013	115	12.6 (0.9)	45	10.7 (1)		2.04 [1.62, 2.45]



Analysis 7.3. Comparison 7 Any psychological versus other psychological (medium term), Outcome 3 Work functioning.

Review: Interventions to improve return to work in depressed people

Comparison: 7 Any psychological versus other psychological (medium term)

Outcome: 3 Work functioning

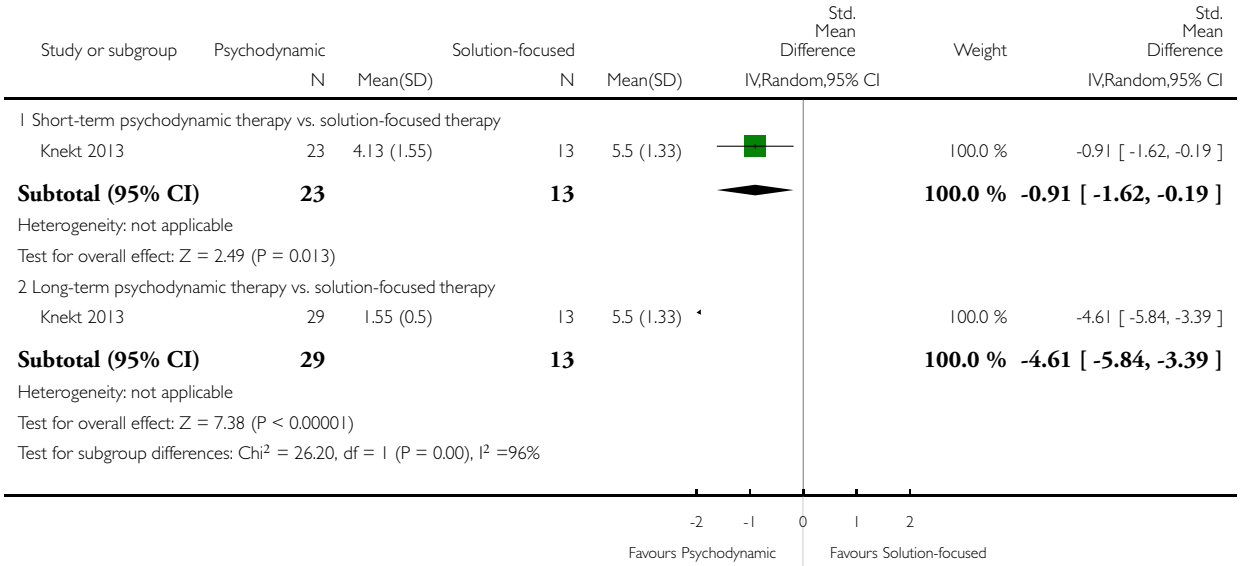
Study or subgroup	Psychodynamic N	Mean(SD)	Solution-focused N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Std. Mean Difference IV,Random,95% CI
1 Short-term psychodynamic therapy vs solution-focused therapy						
Knekt 2013	91	1.91 (0.06)	45	1.95 (0.06)		-0.66 [-1.03, -0.30]
2 Long-term psychodynamic therapy vs solution-focused therapy						
Knekt 2013	115	2.01 (0.06)	45	1.95 (0.06)		1.00 [0.63, 1.36]

Analysis 8.1. Comparison 8 Any psychological versus other psychological (long term), Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 8 Any psychological versus other psychological (long term)

Outcome: 1 Days of sickness absence



Analysis 8.2. Comparison 8 Any psychological versus other psychological (long term), Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 8 Any psychological versus other psychological (long term)

Outcome: 2 Depressive symptoms

Study or subgroup	Psychodynamic N	Mean(SD)	Solution-focused N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Short-term psychodynamic therapy vs. solution-focused therapy							
Knekt 2013	79	8.5 (0.9)	39	9.7 (1)		50.4 %	-1.28 [-1.69, -0.86]
Subtotal (95% CI)	**79**		**39**			**50.4 %**	**-1.28 [-1.69, -0.86]**
Heterogeneity: not applicable							
Test for overall effect: Z = 5.99 (P < 0.00001)							
2 Long-term psychodynamic therapy vs. solution-focused therapy							
Knekt 2013	106	7.6 (0.8)	39	9.7 (1)		49.6 %	-2.44 [-2.90, -1.97]
Subtotal (95% CI)	**106**		**39**			**49.6 %**	**-2.44 [-2.90, -1.97]**
Heterogeneity: not applicable							
Test for overall effect: Z = 10.28 (P < 0.00001)							
Total (95% CI)	**185**		**78**			**100.0 %**	**-1.85 [-2.99, -0.72]**
Heterogeneity: Tau^2 = 0.62; Chi^2 = 13.23, df = 1 (P = 0.00028); I^2 =92%							
Test for overall effect: Z = 3.20 (P = 0.0014)							
Test for subgroup differences: Chi^2 = 13.23, df = 1 (P = 0.00), I^2 =92%							

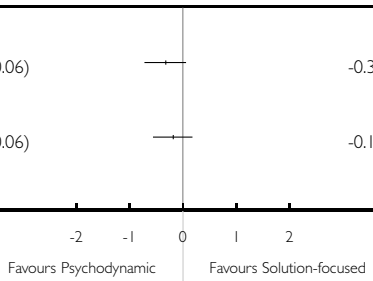
Analysis 8.3. Comparison 8 Any psychological versus other psychological (long term), Outcome 3 Work functioning.

Review: Interventions to improve return to work in depressed people

Comparison: 8 Any psychological versus other psychological (long term)

Outcome: 3 Work functioning

Study or subgroup	Psychodynamic N	Mean(SD)	Solution-focused N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Std. Mean Difference IV,Random,95% CI
1 Short-term psychodynamic therapy vs. solution-focused therapy						
Knekt 2013	79	1.8 (0.06)	39	1.82 (0.06)		-0.33 [-0.72, 0.05]
2 Long-term psychodynamic therapy vs. solution-focused therapy						
Knekt 2013	106	1.81 (0.05)	39	1.82 (0.06)		-0.19 [-0.56, 0.18]



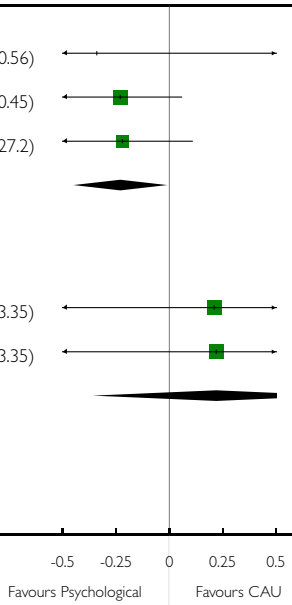
Analysis 9.1. Comparison 9 Any psychological versus no intervention or care as usual, Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 9 Any psychological versus no intervention or care as usual

Outcome: 1 Days of sickness absence

Study or subgroup	Psychological N	Mean(SD)	CAU N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Online/telephone CBT vs. CAU							
Bee 2010	5	3.85 (1.76)	7	4.29 (0.56)		3.5 %	-0.34 [-1.50, 0.82]
McCrone 2004	88	3.84 (19.01)	86	11.22 (40.45)		53.5 %	-0.23 [-0.53, 0.06]
Hollinghurst 2010	74	7.6 (17.9)	66	12.7 (27.2)		42.9 %	-0.22 [-0.56, 0.11]
Subtotal (95% CI)	**167**		**159**			**100.0 %**	**-0.23 [-0.45, -0.01]**
Heterogeneity: Tau² = 0.0; Chi² = 0.04, df = 2 (P = 0.98); I² =0.0%							
Test for overall effect: Z = 2.09 (P = 0.037)							
2 CMHN vs. usual GP care							
Kendrick 2005	20	17.6 (35.07)	8	10.93 (13.35)		49.1 %	0.21 [-0.61, 1.03]
Kendrick 2005	23	17.7 (33.1)	8	10.93 (13.35)		50.9 %	0.22 [-0.58, 1.03]
Subtotal (95% CI)	**43**		**16**			**100.0 %**	**0.22 [-0.36, 0.79]**
Heterogeneity: Tau² = 0.0; Chi² = 0.00, df = 1 (P = 0.98); I² =0.0%							
Test for overall effect: Z = 0.74 (P = 0.46)							
Test for subgroup differences: Chi² = 2.05, df = 1 (P = 0.15), I² =51%							

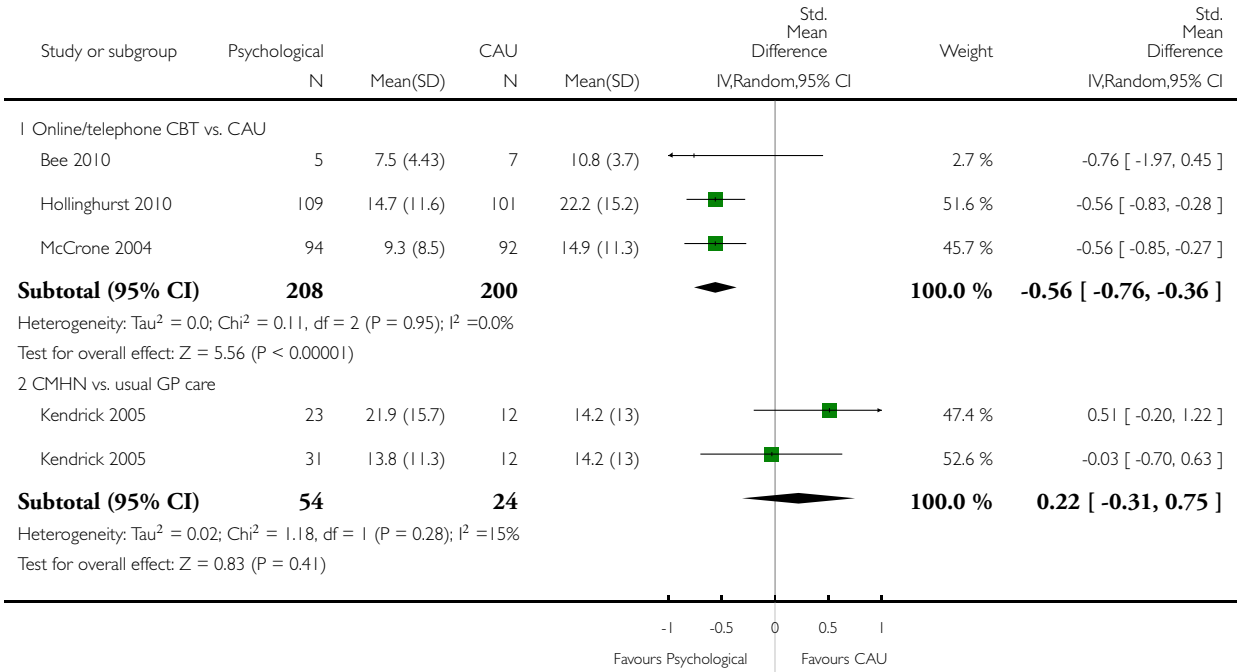


Analysis 9.2. Comparison 9 Any psychological versus no intervention or care as usual, Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 9 Any psychological versus no intervention or care as usual

Outcome: 2 Depressive symptoms

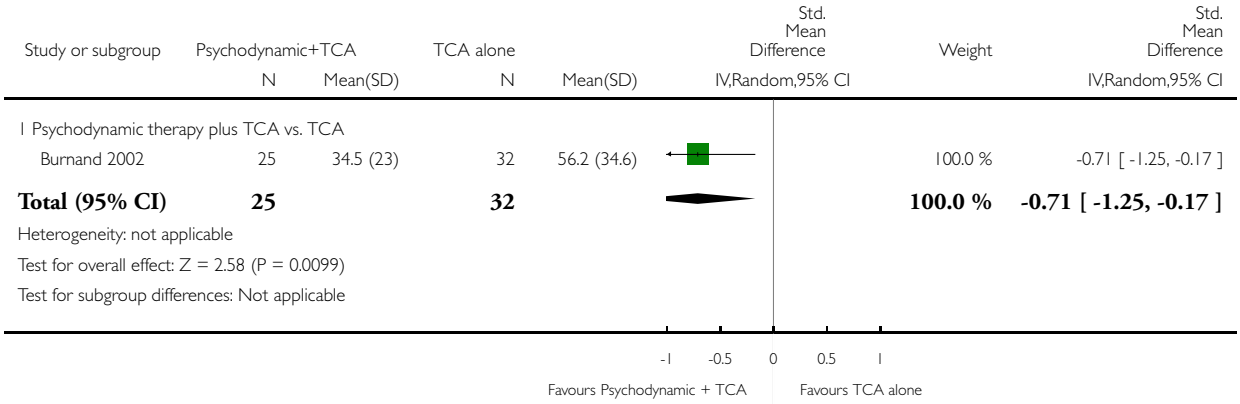


Analysis 10.1. Comparison 10 Psychological combined with antidepressant medication versus antidepressant medication alone, Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 10 Psychological combined with antidepressant medication versus antidepressant medication alone

Outcome: 1 Days of sickness absence

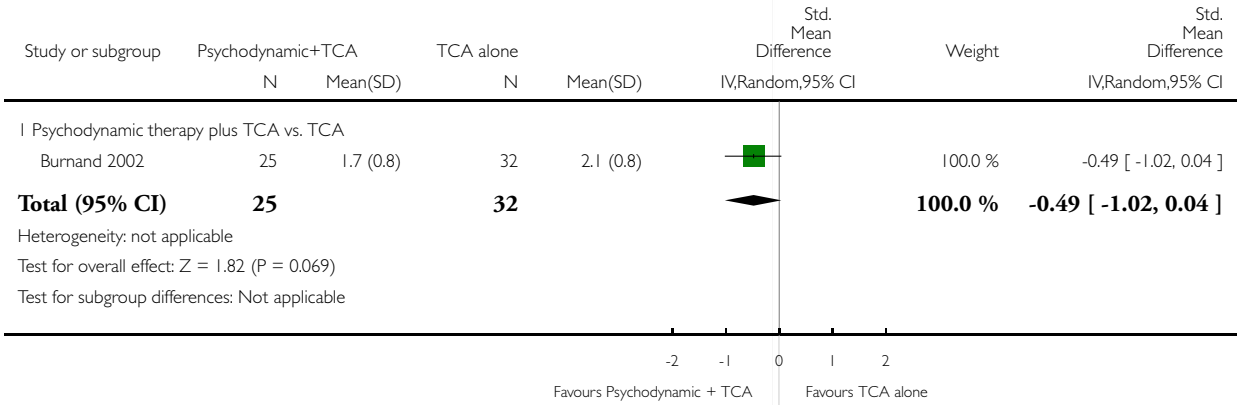


Analysis 10.2. Comparison 10 Psychological combined with antidepressant medication versus antidepressant medication alone, Outcome 2 Work functioning or productivity.

Review: Interventions to improve return to work in depressed people

Comparison: 10 Psychological combined with antidepressant medication versus antidepressant medication alone

Outcome: 2 Work functioning or productivity

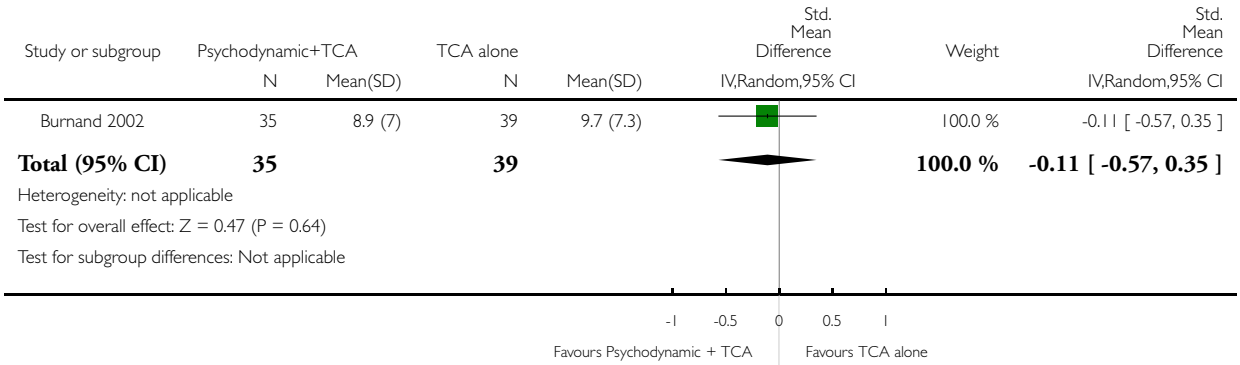


Analysis 10.3. Comparison 10 Psychological combined with antidepressant medication versus antidepressant medication alone, Outcome 3 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 10 Psychological combined with antidepressant medication versus antidepressant medication alone

Outcome: 3 Depressive symptoms



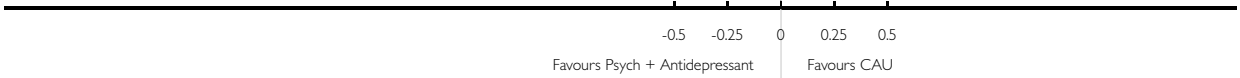
Analysis 11.1. Comparison 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term), Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term)

Outcome: 1 Days of sickness absence

Study or subgroup	Psych.+ Antide-pressant N	Mean(SD)	CAU N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Enhanced primary care vs. CAU							
Schoenbaum 2001	593	1.2 (3.94)	288	1.2 (3.94)		89.9 %	0.0 [-0.14, 0.14]
Simon 1998	48	2.02 (7.82)	40	4.02 (14.46)		10.1 %	-0.18 [-0.60, 0.25]
Subtotal (95% CI)	**641**		**328**			**100.0 %**	**-0.02 [-0.15, 0.12]**
Heterogeneity: $Tau^2 = 0.0$; $Chi^2 = 0.60$, df = 1 (P = 0.44); $I^2 = 0.0\%$							
Test for overall effect: Z = 0.26 (P = 0.80)							
2 Telephone outreach and care management program vs. CAU							
Wang 2007	304	-0.5375 (1.675)	300	-0.19 (1.7125)		100.0 %	-0.21 [-0.37, -0.05]
Subtotal (95% CI)	**304**		**300**			**100.0 %**	**-0.21 [-0.37, -0.05]**
Heterogeneity: not applicable							
Test for overall effect: Z = 2.53 (P = 0.011)							
Test for subgroup differences: $Chi^2 = 3.15$, df = 1 (P = 0.08), $I^2 = 68\%$							

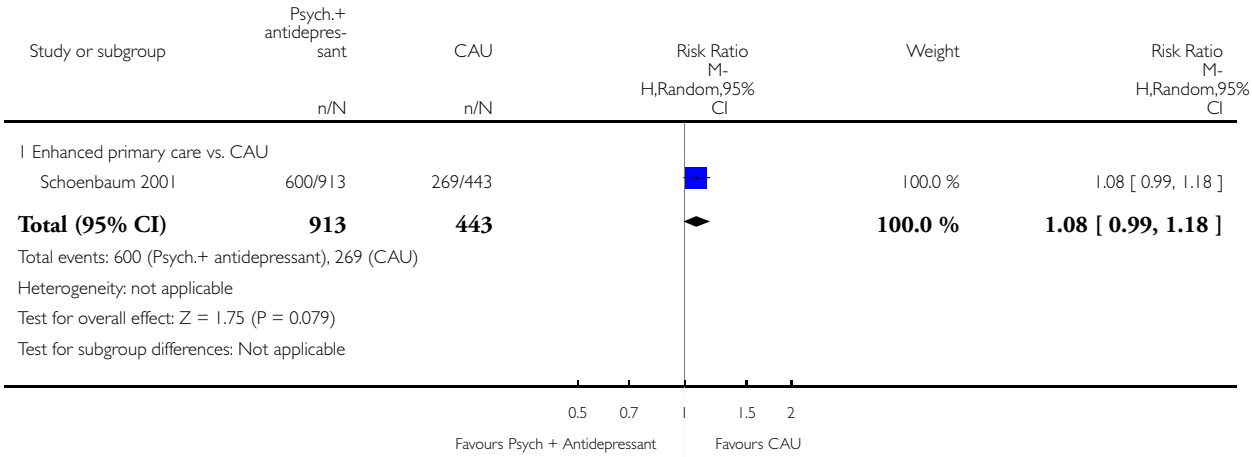


Analysis 11.2. Comparison 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term), Outcome 2 Employment status.

Review: Interventions to improve return to work in depressed people

Comparison: 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term)

Outcome: 2 Employment status



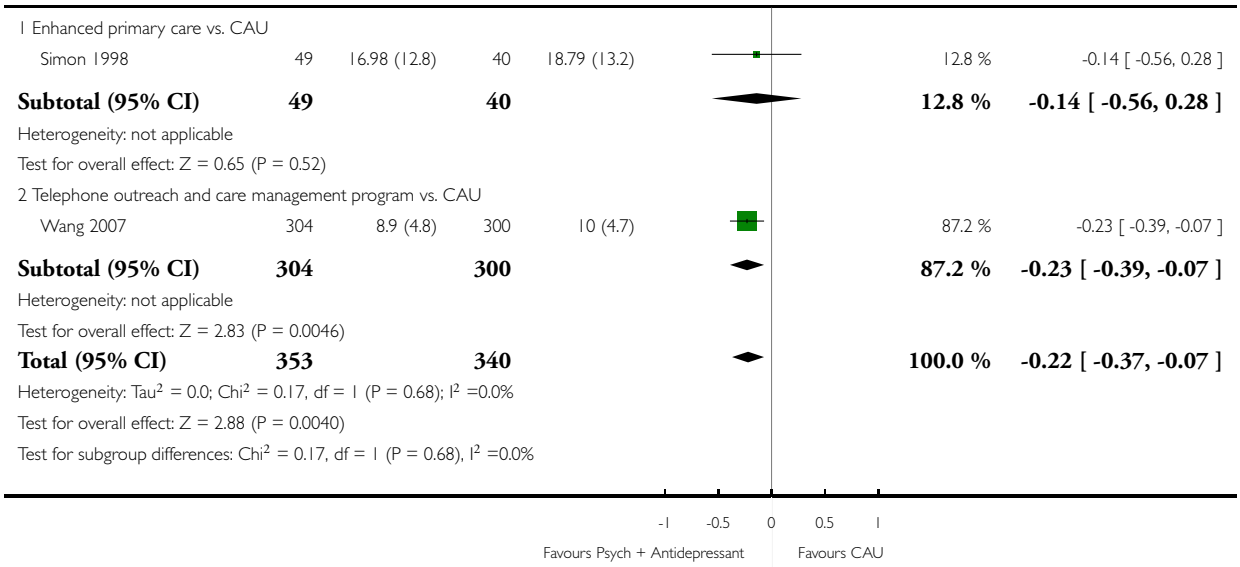
Analysis 11.3. Comparison 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term), Outcome 3 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term)

Outcome: 3 Depressive symptoms

Study or subgroup	Psych.+ antidepressant N	Mean(SD)	CAU N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Enhanced primary care vs. CAU							
Simon 1998	49	16.98 (12.8)	40	18.79 (13.2)		12.8 %	-0.14 [-0.56, 0.28]
Subtotal (95% CI)	**49**		**40**			**12.8 %**	**-0.14 [-0.56, 0.28]**
Heterogeneity: not applicable							
Test for overall effect: Z = 0.65 (P = 0.52)							
2 Telephone outreach and care management program vs. CAU							
Wang 2007	304	8.9 (4.8)	300	10 (4.7)		87.2 %	-0.23 [-0.39, -0.07]
Subtotal (95% CI)	**304**		**300**			**87.2 %**	**-0.23 [-0.39, -0.07]**
Heterogeneity: not applicable							
Test for overall effect: Z = 2.83 (P = 0.0046)							
Total (95% CI)	**353**		**340**			**100.0 %**	**-0.22 [-0.37, -0.07]**
Heterogeneity: Tau^2 = 0.0; Chi^2 = 0.17, df = 1 (P = 0.68); I^2 =0.0%							
Test for overall effect: Z = 2.88 (P = 0.0040)							
Test for subgroup differences: Chi^2 = 0.17, df = 1 (P = 0.68), I^2 =0.0%							

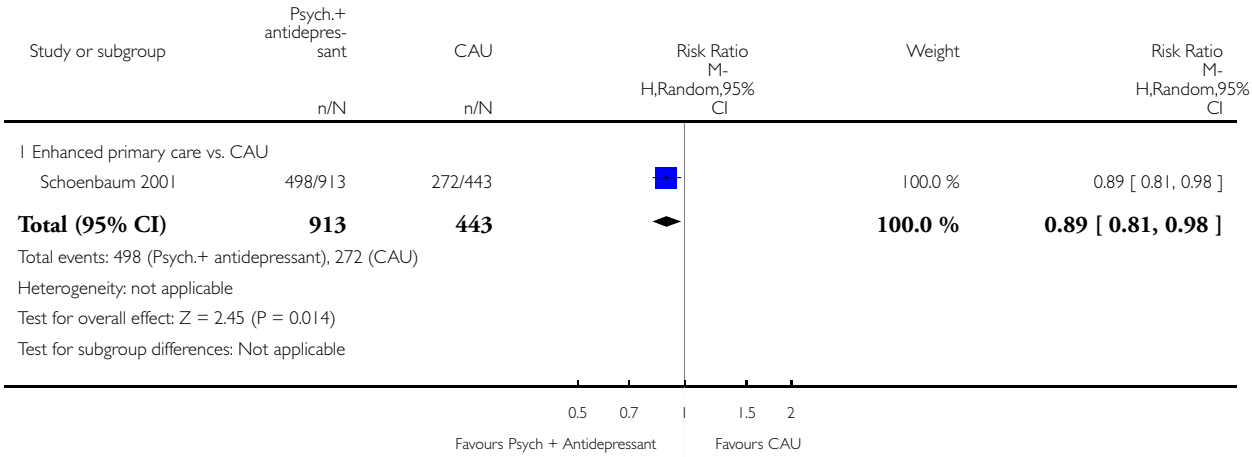


Analysis 11.4. Comparison 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term), Outcome 4 Depressed yes/no.

Review: Interventions to improve return to work in depressed people

Comparison: 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term)

Outcome: 4 Depressed yes/no



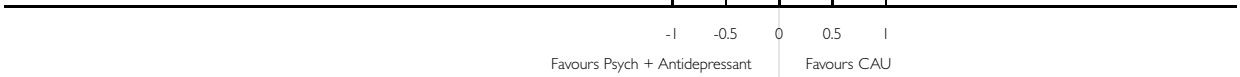
Analysis 11.5. Comparison 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term), Outcome 5 Work functioning.

Review: Interventions to improve return to work in depressed people

Comparison: 11 Psychological combined with antidepressant medication versus no intervention or usual care (medium term)

Outcome: 5 Work functioning

Study or subgroup	Psych.+ antidepres-sant		CAU		Std. Mean Difference	Weight	Std. Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Random,95% CI		IV,Random,95% CI
1 Telephone outreach and care management program vs. CAU							
Wang 2007	304	0.8 (0.2)	300	0.7 (0.2)		100.0 %	0.50 [0.34, 0.66]
Total (95% CI)	**304**		**300**			**100.0 %**	**0.50 [0.34, 0.66]**
Heterogeneity: not applicable							
Test for overall effect: Z = 6.04 (P < 0.00001)							
Test for subgroup differences: Not applicable							

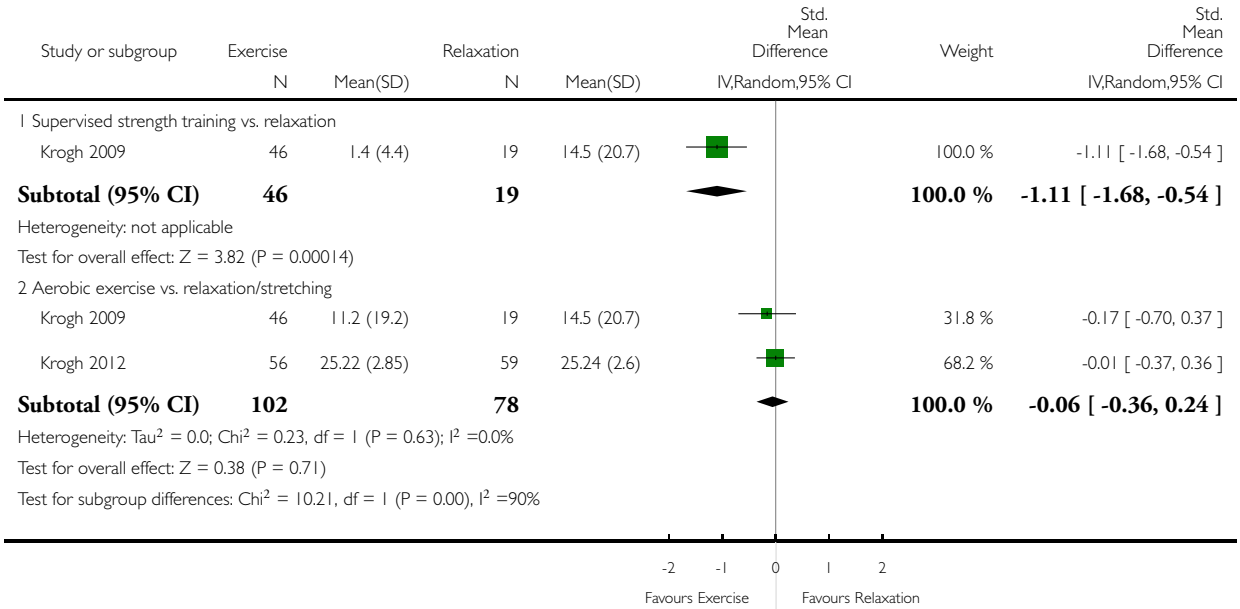


Analysis 12.1. Comparison 12 Exercise intervention versus no intervention or care as usual, Outcome 1 Days of sickness absence.

Review: Interventions to improve return to work in depressed people

Comparison: 12 Exercise intervention versus no intervention or care as usual

Outcome: 1 Days of sickness absence

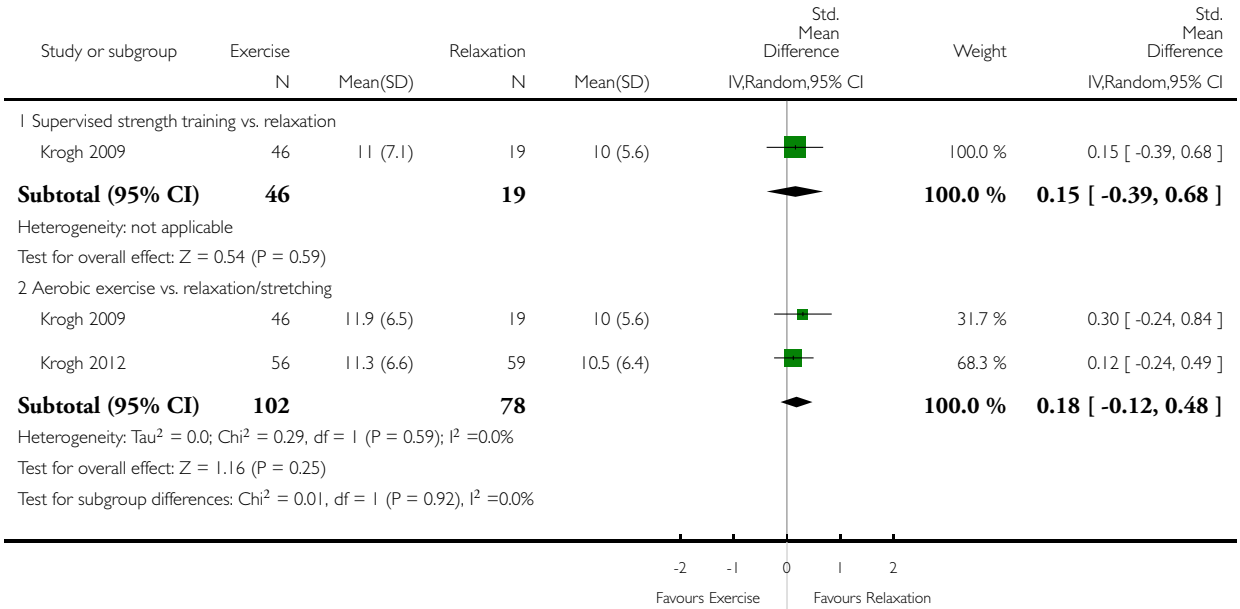


Analysis 12.2. Comparison 12 Exercise intervention versus no intervention or care as usual, Outcome 2 Depressive symptoms.

Review: Interventions to improve return to work in depressed people

Comparison: 12 Exercise intervention versus no intervention or care as usual

Outcome: 2 Depressive symptoms



ADDITIONAL TABLES

Table 1. Quality of the evidence (GRADE)

Comparison	Studies in comparison	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication bias	Quality of the evidence
Work-directed							
Work-directed + clinical intervention versus clinical (medium)	Hees 2013; Lerner 2012; Schene 2006	No: Majority low risk of bias	No: I^2 < 50%	No	Yes: N < -400 Downgrade: -1	Undetected	Moderate

Table 1. Quality of the evidence (GRADE) (*Continued*)

Work-directed + clinical intervention versus clinical (long)	Hees 2013; Schene 2006	No: Majority low risk of bias	No: I^2 < 50%	No	Yes: N < -400 Downgrade: -1	Undetected	Moderate
Work-directed + clinical versus work-directed	Vlasveld 2013	No: study with low risk of bias	N.a.	No	Yes: Single study Downgrade: -1	Undetected	Moderate
Work-directed versus work-directed	Noordik 2013	Yes: very serious Downgrade -2	N.a.	No	Yes: wide CI Downgrade: -1	Undetected	Very low
Clinical, medication							
Medication: SSRI versus SNRI	Fernandez 2005; Romeo 2004; Wade 2008	Fernandez: low Romeo: high Wade: high	Yes: 83%, pooling not feasible	No	Yes: Single studies Downgrade: -1	Undetected	Fernandez: moderate Romeo: Low Wade: Low
SSRI versus TCA	Miller 1998	Yes: very serious Downgrade -2	N.a.	No	No	Undetected	Low
SSRI versus SSRI	Fantino 2007	No: study with low risk of bias	N.a.	No	Yes: N< -400 Downgrade: -1	Undetected	Moderate
TCA or MAO versus placebo	Agosti 1991	Yes: very serious Downgrade -2	N.a.	No	Yes: N < -400 Downgrade: -1	Undetected	Very low
Clinical, psychological							
Any psych versus other psych (medium)	Knekt 2013 (three arms)	Yes: serious Downgrade -1	Yes: 99% pooling not feasible	No	Yes: single study arms Downgrade -1	Undetected	Knekt I: Low Knekt II: Low
Any psych versus other psych	Knekt 2013 (three arms)	Yes: serious Downgrade -1	Yes: 99% pooling not feasible	No	Yes: single study arms Downgrade -1	Undetected	Knekt I: Low Knekt II: Low

Table 1. Quality of the evidence (GRADE) (*Continued*)

(medium)							
Any psych versus CAU	Bee 2010; Hollinghurst 2010; McCrone 2004	No: Majority low risk of bias	No: I^2 < 50%	No	Yes: N < -400 Downgrade: -1	Undetected	Moderate
CMHN versus CAU	Kendrick 2005 (three arms)	Yes: serious Downgrade -1	No: I^2 < 50%	No	Yes: wide CI Downgrade: -1	Undetected	Low
Clinical, psychological and medication							
Psych + med versus medicine	Burnand 2002	Yes: very serious Downgrade -2	N.a.	No	Yes: N < -400 Downgrade: -1	Undetected	Very low
Enhanced primary care versus CAU	Rost 2004; Schoenbaum 2001; Simon 1998	Yes: very serious Downgrade -2	No: I^2 < 50%	No	No	Undetected	Low
Telephone outreach versus CAU	Wang 2007	No: study with low risk of bias	N.a.	No	No (n > 400 * CI not wide)	Undetected	High
Clinical, exercise							
Strength versus relax	Krogh 2009	Yes: serious Downgrade -1	N.a.	No	Yes: N < -400 Downgrade: -1	Undetected	Low
Aerobic versus relax or stretching	Krogh 2009; Krogh 2012	No: Majority low risk of bias	No: I^2 < 50%	No	Yes: N < -400 Downgrade: -1	Undetected	Moderate

APPENDICES

Appendix 1. Search strategy update 2006 to 2014

MEDLINE (via Ovid)

1. exp Depressive Disorder/
2. exp DEPRESSION/
3. exp Adjustment Disorders/
4. exp Mood Disorders/
5. exp Affective Symptoms/
6. 1 or 2 or 3 or 4 or 5
7. exp Occupational Therapy/
8. exp Occupational Diseases/
9. exp Occupational Medicine/
10. exp Disability Evaluation/
11. exp WORK/
12. return to work.mp.
13. occupational therap$.mp.
14. occupational intervention$.mp.
15. supported employment.mp.
16. employment.mp.
17. vocational rehabilitation.mp.
18. work capacity evaluation.mp.
19. vocational guidance.mp.
20. absenteeism.mp.
21. occupational health services.mp.
22. occupational health.mp.
23. unemployed.mp.
24. employed.mp.
25. unemployment.mp.
26. sick leave.mp.
27. sick$ absence.mp.
28. retirement.mp.
29. disability pension.mp.
30. occupation$.mp.
31. job.mp.
32. vocational.mp.
33. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
34. randomized controlled trial.pt. OR randomized.mp. OR placebo.mp.
35. 6 and 33 and 34

EMBASE (via Ovid)

1. exp depression/
2. exp mood disorder/
3. exp adjustment disorder/
4. 1 or 2 or 3
5. occupational therapy.mp.
6. occupational disease.mp.
7. occupational medicine.mp.

8. employment.mp.
9. vocational rehabilitation.mp.
10. work capacity.mp.
11. vocational guidance.mp.
12. absenteeism.mp.
13. occupational health service.mp.
14. occupational health.mp.
15. unemployment.mp.
16. retirement.mp.
17. occupation.mp.
18. vocation.mp.
19. disability evaluation.mp.
20. return to work.mp.
21. occupational intervention$.mp.
22. supported employment.mp.
23. unemployed.mp.
24. employed.mp.
25. sick leave.mp.
26. sick$ absence.mp.
27. disability pension.mp.
28. job.mp.
29. vocational.mp.
30. exp work/
31. (disability adj (work or occupation$ or vocation$ or job)).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]
32. 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31
33. Random:.tw. OR placebo:.mp. OR double-blind:.tw.
34. 4 and 32 and 33

PsycINFO (via Ovid)

1. exp Affective Disorders/
2. exp Major Depression/
3. "depression (emotion)".mp.
4. exp Dysthymic Disorder/
5. Neurotic Depressive Reaction.mp.
6. exp Reactive Depression/
7. exp Recurrent Depression/
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. exp Disability Evaluation/
10. exp Employability/
11. exp Employee Leave Benefits/
12. exp Job Satisfaction/
13. exp Occupational Guidance/
14. exp Vocational Rehabilitation/
15. exp Disability Management/
16. exp Employee Absenteeism/
17. exp Occupational Status/
18. exp Occupational Stress/
19. exp Occupational Therapy/
20. exp Reemployment/
21. exp Work Related Illnesses/

22. return to work.ti,ab,tc.
23. occupational therap*.ti,ab,tc.
24. occupational intervention*.ti,ab,tc.
25. Supported employment.ti,ab,tc.
26. employment.ti,ab,tc.
27. vocational rehabilitation.ti,ab,tc.
28. work capacity evaluation.ti,ab,tc.
29. vocational guidance.ti,ab,tc.
30. Absenteeism.ti,ab,tc.
31. Occupational health services.ti,ab,tc.
32. Occupational health.ti,ab,tc.
33. Unemployed.ti,ab,tc.
34. Employed.ti,ab,tc.
35. Unemployment.ti,ab,tc.
36. Sick leave.ti,ab,tc.
37. Sick* absence.ti,ab,tc.
38. Retirement.ti,ab,tc.
39. Disability pension.ti,ab,tc.
40. Occupation*.ti,ab,tc.
41. Job.ti,ab,tc.
42. Vocational.ti,ab,tc.
43. 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42
44. random*.ti,ab,tc.
45. ((singl* or doubl* or trebl* or tripl*) adj26 (blind* or dummy or mask*)).ti,ab,tc.
46. placebo*.ti,ab,tc.
47. Crossover.ti,ab,tc.
48. Assign*.ti,ab,tc.
49. Allocat*.ti,ab,tc.
50. ((clin* or control* or compare* or evaluat* or prospective*) adj26 (trial* or studi* or study)).ti,ab,tc.
51. exp Placebo/
52. exp Treatment Effectiveness Evaluation/
53. exp Mental Health Program Evaluation/
54. exp Experimental Design/
55. (assign* or crossover or placebo* or ((singl* or doubl* or trebl* or tripl*) adj26 (blind* or dummy or mask*))).ti,ab,tc. or explode experimental design/ or random*.ti,ab,tc. or explode mental health program evaluation/ or explode treatment effectiveness evaluation/ or explode placebo/ or ((clin* or control* or compare* or evaluat* or prospective*) adj26 (trial* or studi* or study)).ti,ab,tc. or allocat*.ti,ab,tc.
56. Animal.po.
57. (human or inpatient or outpatient).po.
58. ((human or inpatient or outpatient) and animal).po.
59. (56 not 58)
60. (55 not 59)
61. 8 and 43 and 60

CINAHL (via EBSCOhost)

1. "Depression"
2. (MH "Affective Disorders+")
3. (MH "Affective Symptoms+")
4. (MH "Adjustment Disorders+")
5. (MH "Neurotic Disorders+")
6. S1 OR S2 OR S3 OR S4 OR S5

7. (MH “Job Performance”)
8. (MH “Job Re-Entry”)
9. (MH “Employment+”)
10. (MH “Occupational Health+”)
11. (MH “Rehabilitation, Vocational+”)
12. (MH “Sick Leave”)
13. (MH “Work”)
14. (MH “Disability Evaluation+”)
15. (MH “Occupational Therapy+”)
16. TI Return to work OR AB Return to work OR SU Return to work
17. TI Occupational therap* OR AB Occupational therap* OR SU Occupational therap*
18. TI Occupational intervention* OR AB Occupational intervention* OR SU Occupational intervention*
19. TI Supported employment OR AB Supported employment OR SU Supported employment
20. TI employment OR AB Employment OR SU Employment
21. TI vocational rehabilitation OR AB vocational rehabilitation OR SU vocational rehabilitation
22. TI Work capacity evaluation OR AB Work capacity evaluation OR SU Work capacity evaluation
23. TI vocational guidance OR AB vocational guidance OR SU vocational guidance
24. TI absenteeism OR AB absenteeism OR SU absenteeism
25. TI occupational health services OR AB occupational health services OR SU occupational health services
26. TI occupational health OR AB occupational health OR SU occupational health
27. TI unemployed OR AB unemployed OR SU unemployed
28. TI employed OR AB employed OR SU employed
29. TI unemployment OR AB unemployment OR SU unemployment
30. TI Sick leave OR AB sick leave OR SU sick leave
31. TI Sick* absence OR AB sick* absence OR SU sick* absence
32. TI retirement OR AB retirement OR SU retirement
33. TI Disability pension OR AB Disability pension OR SU Disability pension
34. TI Occupation* OR AB Occupation* OR SU Occupation
35. TI Job OR AB Job OR SU Job
36. TI vocational OR AB vocational OR SU vocational
37. S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36
38. PT clinical trial
39. (MH “Clinical Trials+”)
40. TI (clin* N24 trial*) OR AB (clin* N24 trial*)
41. TI (((singl* or doubl8 or tripl* or trebl*) N24 (blind* or mask* or dummy*))) OR AB (((singl* or doubl8 or tripl* or trebl*) N24 (blind* or mask* or dummy*))) OR SU (((singl* or doubl8 or tripl* or trebl*) N24 (blind* or mask* or dummy*)))
42. (MH “Placebos”)
43. TI placebo* OR AB placebo*
44. TI random* OR AB random*
45. (MH “Evaluation Research+”)
46. (MH “Prospective Studies”)
47. TI ((control* or prospectiv* or volunteer*)) OR AB ((control* or prospectiv* or volunteer*))
48. S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47
49. S6 AND S37 AND S48

CENTRAL (The Cochrane Library)

#1 depressive disorder
#2 depression
#3 Mood Disorders
#4 #1 or #2 or #3
#5 Occupational Therapy

#6 Occupational Diseases
#7 Occupational Medicine
#8 return to work
#9 occupational intervention$
#10 absenteeism
#11 occupational health services
#12 occupational health
#13 disability pension
#14 sick leave
#15 sick$ absence
#16 #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15
#17 #4 and #16 in Trials

Appendix 2. Search strategy up until 2006

First, we searched two Cochrane Depression Anxiety Neurosis Group specialised registers (study-based and reference-based) to identify all potentially eligible studies. We used both work terms as well as terms relating to depression:
CCDANCTR-Studies (searched on 2/8/2006)
Diagnosis = Depress* or Dysthymi* or "Mood Disorder*" or "Affective Disorder" or "Affective Symptoms"
and
Setting = work*
or
Outcomes = Work* or employ* or vocation* or occupat* or "sick days" or "Sick Leave" or "Sick Absence" or "Time Off"
CCDANCTR-References (searched on 2/8/2006)
Keyword = Depress* or Dysthymi* or "Mood Disorder*" or "Affective Disorder" or "Affective Symptoms"
and
Free-text = ("occupational" and (intervention* or therap* or treatment*)) and (work* or employe* or employment* or vocation* or "sick leave" or disabil* or absentee*)
Second, we searched the following electronic databases up to August 2006: CENTRAL, MEDLINE, EMBASE, CINAHL, PsycINFO, OSH-ROM (Occupational Safety and Health; all databases except for MEDLINE), NHS-EED (1994 to August 2006), and the Database of Abstracts of Reviews of Effectiveness (DARE).
In MEDLINE, PsycINFO, EMBASE and CINAHL and OSH-ROM we used three types of terms: depression-related words (see CCDAN search strategy) combined with work-related words and database-specific methodological filters terms (see CCDAN search strategy).

WHAT'S NEW

Date	Event	Description
6 June 2014	New citation required and conclusions have changed	Full update. This updated review includes 12 new studies with 3440 new participants (added to the 11 studies with 2556 participants of the former version). We have modified the names of the interventions in the comparisons: we now include work-directed and clinical interventions, while in the 2008 version clinical interventions were under worker-directed interventions. In the update, we refrained from handsearching journals as this strategy did not yield additional studies in the 2008 version. We have re-assessed all studies that we originally

(*Continued*)

		included to be able to use the GRADE method. Two new authors have joined the review team: Babs Faber and Hiske Hees

HISTORY

Date	Event	Description
2 November 2008	Amended	Converted to new review format.
20 November 2007	New citation required and conclusions have changed	Substantive amendment

CONTRIBUTIONS OF AUTHORS

Original review

KN wrote the initial draft of the protocol and will write subsequent drafts of the protocol and review. She and AN designed and conducted the search strategy. AV, UB, CF, AN, and JV contributed to the draft version of the protocol and contributed to subsequent versions and revisions of the protocol and review. KN, AV, and UB included eligible studies. UB and CF conducted the quality assessment of eligible studies. KN and AN extracted the data from the original studies. KN, CF, and JV conducted the data synthesis.

Update 2014

BF adapted the search strategy and conducted the searches. BF, KN, CF, UB, and AV checked resulting studies for eligibility. BF, KN, AN, AV CF, HH, and UB conducted data extraction. BF, KN, AN, AV, CH, HH, UB, and JV assessed included studies for risk of bias. BF, KN, and JV ran the analyses. KN wrote the draft of the updated review and all others commented on this draft. JV acted as an advisor on the whole review process and several specific topics such as meaningful comparisons, GRADE, and meta-analysis.

DECLARATIONS OF INTEREST

Karen Nieuwenhuijsen was an author of one of the included studies: Noordik 2013.

Babs Faber: none known.

Jos Verbeek: none known.

Angela Neumeyer-Gromen: none known.

Hiske Hees was an author of one of the included studies: Hees 2013.

Arco Verhoeven: none known.

Christina van der Feltz-Cornelis was an author of one of the included studies: Vlasveld 2013. Her employer received an unrestricted grant from Eli Lilly for an investigator-initiated trial on depression and pain. She also received payment from Benecke for speaking at a symposium on chronic pain. She has received royalties from various publishers on her books on psychiatry.

Ute Bültmann: none known.

None of the authors assessed studies they were authors of for eligibility or risk of bias.

SOURCES OF SUPPORT

Internal sources

- Coronel Institute of Occupational Health, Netherlands.

Salary for Karen Nieuwenhuijsen and Babs Faber

- Trimbos Instituut - Netherlands Institute of Mental Health and Addiction, Netherlands.

Salary for Christina van der Feltz-Cornelis

- Federal Institute for Occupational Safety and Health, Germany.

Salary for Angela Neumeyer-Gromen

- Finnish Institute of Occupational Health, Finland.

Salary for Jos Verbeek

- University Medical Center Groningen, Netherlands.

Salary for Ute Bültmann

- Dutch Research Center for Insurance Medicine, Netherlands.

Support and training for authors

External sources

- KIS programme, Ministry of Social Affairs and Employment, Netherlands.

A small grant to Karen Nieuwenhuijsen to help her finish the first version of this review

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

In order to reflect the latest guidance available in the *Cochrane Handbook for Systematic Teviews of Interventions*, we used the GRADE approach. In the former version of the protocol and the published review, we used the Downs and Black checklist to assess quality, while in this update we used the Cochrane Collaboration's risk of bias tool. Also, we no longer formally tested heterogeneity but rather assessed the I^2 statistic. Furthermore, our search strategy was simplified and we no longer handsearched journals as these were indexed in MEDLINE and did not yield additional studies. Instead of searching the CCDAN registers, we now directly searched CENTRAL.

INDEX TERMS

Medical Subject Headings (MeSH)

*Absenteeism; *Occupational Health; Antidepressive Agents [therapeutic use]; Cognitive Therapy; Depression [*therapy]; Depressive Disorder, Major [*therapy]; Muscle Stretching Exercises; Randomized Controlled Trials as Topic; Return to Work [*psychology]; Sick Leave

MeSH check words

Adult; Humans